The role of predictive processing in conscious access and regularity learning across sensory domains

Acer Yu-Chan. Chang

Dissertation submitted for the degree of

doctor of philosophy

School of Engineering and Informatics

Department of Informatics

University of Sussex

December 2016

Summary

To increase fitness for survival, organisms not only passively react to environmental changes but also actively predict future events to prepare for potential hazards within their environment. Accumulating evidence indicates that the human brain is a remarkable predictive machine which constantly models causal relationships and predicts future events. This 'predictive processing' framework, a prediction-based form of Bayesian inference, states that the brain continuously generates and updates predictions about incoming sensory signals. This framework has been showing notable explanatory power in understanding the mechanisms behind both human behaviour and neurophysiological data and elegantly specifies the underlying computational principles of the neural system. However, even though predictive processing has the potential to provide a unified theory of the brain (Karl Friston, 2010), we still have a limited understanding about fundamental aspects of this model, such as how it deals with different types of information, learns statistical regularities and perhaps most fundamentally of all what its relationship to conscious experience is. This thesis aims to investigate the major gaps in our current understanding of the predictive processing framework via a series of studies. Study 1 investigated the fundamental relationship between unconscious statistical inference reflected by predictive processing and conscious access. It demonstrated that predictions that are in line with sensory evidence accelerate conscious access. Study 2 investigated how low level information within the sensory hierarchy is dealt with by predictive processing and regularity learning mechanisms through "perceptual echo" in which the cross-correlation between a sequence of randomly fluctuating luminance values and occipital electrophysiological (EEG) signals exhibits a long-lasting periodic (~100ms cycle) reverberation of the input stimulus. This study identified a new form of regularity learning and the results demonstrate that the perceptual echo may reflect an iterative learning process, governed by predictive processing. Study 3 investigated how supra-modal predictive processing is capable of

learning regularities of temporal duration and also temporal predictions about future events. This study revealed a supramodal temporal prediction mechanism which processes auditory and visual temporal information and integrates information from the duration and rhythmic structures of events. Together these studies provide a global picture of predictive processing and regularity learning across differing types of predictive information.

Table of Content

Acknowledgement

I would like to thank my supervisor Professor Anil Seth and Dr. Ryota Kanai. Professor Anil Seth always proposes most critical and high-standard challenges to me. Without him, my whole work would have become just random pieces without any sensible structure. Dr. Ryota Kanai always provides valuable experimental and research suggestions. I am very lucky that I can join University of Sussex at the same time with him.

Then, I need to give my special thanks to my colleague Dr. David Schwartzman. I really can't imagine how I can complete my PhD without him. In terms of research, he helped me with thesis contents, structures, and wording again and again. I believe this almost can be his second PhD thesis. Furthermore, he took care my mental health at the final stage of my PhD. His constant mental support is the main reason that I can keep moving toward to the end of my PhD program.

I very appreciate Dr. Warrick Roseboom's involvement in my final project. I can say the final project is the product of our intellectual interaction. We have complementary skill sets and always can yield good research ideas. I am very happy to work with him.

I would also like to thank my friends in the lab. Maxine Sherman, Michael Schartner, Georgina Heron, Georg Schauer, and Hielke Prins always give me helpful life and research suggestions. It's great that we can study together here.

Finally, I would like to thank my colleagues, Keisuke Suzuki, Jim Parkinson, Lionel Barnett and Sarah Garfinkel, in Sackler Centre for Consciousness Science for precious technical and intellectual supports.

Chapter 1
Introduction

This Chapter will start with a historical review of predictive processing starting with its early formulations in visual science to the Bayesian brain hypothesis, a theoretical antecedent of predictive processing. It will go on to discuss several key models that have substantially influenced the direction of predictive processing e.g. Friston (2005), Rao and Ballard (1999). This Chapter then moves on to review key aspects of predictive processing, unconscious inference, hierarchical predictive processing and regularity learning. The review reveals that we still have a limited understanding about fundamental aspects of predictive processing, such as how it deals with different types of information, learns statistical regularities and perhaps most fundamentally of all what its relationship to conscious experience is. Finally based on this literature review this Chapter discusses the research questions and motivations for the three studies that comprise this thesis.

1.1 Predictive Processing

Predictive processing – a theoretical framework of perception and brain function, which suggests that the brain continuously generates and updates predictions about incoming sensory signals, represents a general principle of neural functioning. The predictive processing framework has been shown to have a high explanatory power across a wide range of domains including auditory (Baldeweg, 2007; Wacongne et al., 2011), visual (Rao & Ballard, 1999; Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008), sensory-motor (Hickok, Houde, & Rong, 2011) and interoceptive (Seth, Suzuki, & Critchley, 2012) processing. Predictive processing is becoming a powerful paradigm in cognitive neuroscience (Clark, 2012), especially in understanding the computational architecture of perception. Predictive-processing theory (Friston, 2009) proposes

that sensory input is compared with predictions generated by a hierarchically organized predictive model in order to minimize surprise. The original predictive processing model which attracted wide attention within neuroscience was a model of the visual cortex proposed by Rao and Ballard (1999). This was a specific model implemented by a hierarchical network and at the time a lot of researchers within the field of neuroscience recognised the huge potential of predictive coding. This led to a more general instance of the original idea, broadly referred to as predictive processing, which describes the brains attempt to constantly model external causal relationships or statistical regularities and predict sensory events. Although the predictive processing framework has superior explanatory power, its relation to conscious experience and whether it is an universal principle of neural processing (Karl Friston, 2010) are still unknown. This thesis will explore this predictive processing definition of perception and brain function by investigating conscious access, regularity learning, and temporal prediction in a series of studies to gain a deeper understanding of the predictive brain.

1.1.1 Inverse problem and Bayesian brain hypothesis

Our sensory systems are constantly bombarded by noisy and ambiguous input. This presents our sensory systems with the difficult task of tracking, decoding, and forming percepts of relevant input which may have multiple interpretations or causes (S. E. Palmer, 1999). A cause here can be defined as the source that causes sensory inputs. A simple example of a possible percept is shown in Figure 1-1 (a), which appears to be a cube. However, this 2D projection on the retina may be due to multiple possible external objects (Figure 1-1b). To infer the actual cause of percept is an unsolvable problem, commonly referred to as the “inverse problem of perception”. This specifies that to infer the external sensory causes that leads to a particular activation of sensory systems is challenging. To understand how the neural system deals with this problem, von Helmholtz

proposed that perception is a statistical inference process. His idea was that neural systems compute perceptual information from sensory data to construct a probabilistic model of the external world where the ultimate goal of perception is to infer the causes that generate sensory inputs (Dayan, Hinton, Neal, & Zemel, 1995; Karl Friston, 2012; Helmholtz, 1866). In this theory, a percept is not only passively caused by sensory stimulation but involves unconscious inferences, associated sensations, and the incorporation of prior knowledge which biases the computation toward a simple interpretation (Pizlo, 2001).

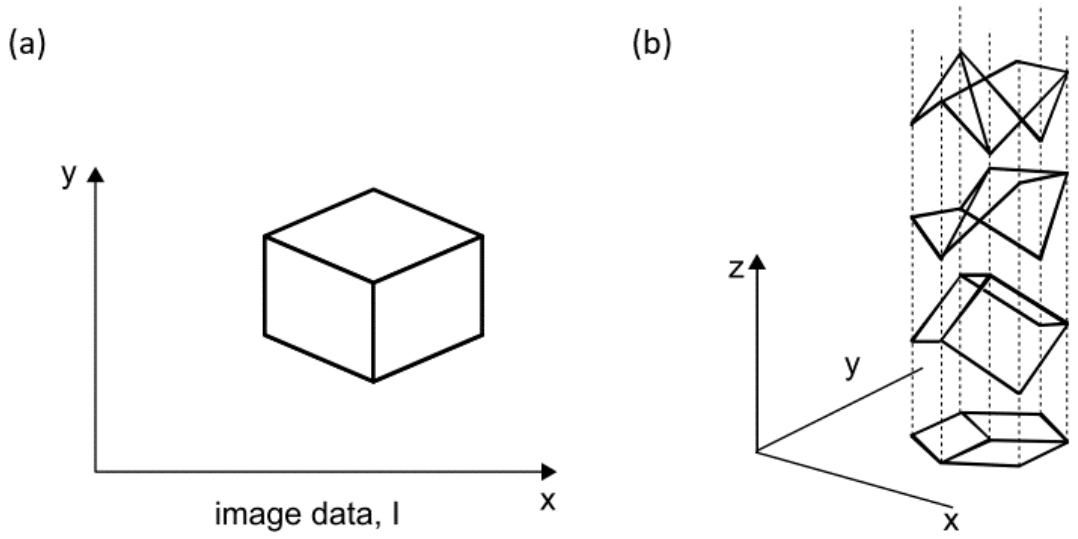

Figure 1-1(a) an simple example of a visual percept (b) multiple possibilities (hidden causes) that can cause the percept (figure taken from Kersten & Yuille, 2003)

Motivated by von Helmholtz, in the past century, accumulating evidence in Psychology and Neuroscience indicates human cognition may process probabilistically. Specifically, the neural system performs Bayesian inference to solve the inverse problem (Friston & Stephan, 2007; Kersten & Schrater, 2002; Knill & Pouget, 2004; Knill & Richards, 1996).

Bayesian inference is a parameter estimation framework that adopts Bayes' theorem to infer states of unknown hidden causes when more information (evidence) becomes available. To

perform this inference, a Bayesian inference process first constructs a prior distribution, a belief probabilistic distribution of all possible hypotheses $p(H)$ based on prior knowledge. When more evidence becomes available the system updates its belief probability distribution (posterior distribution), by updating $p(H)$ to $p(H|Evidence)$ using the following Bayesian rule to integrate prior knowledge and new information:

$$p(H|E) = \frac{p(E|H) \times p(H)}{p(E)} \quad (1\text{-}1)$$

$p(E|H)$ is referred to as the likelihood, which indicates the compatibility of the evidence with the given hypothesis. $p(H)$ is the prior belief of all possible hypotheses. $p(H|E)$ is the posterior distribution indicating the updated belief of all possible hypothesis. $p(E)$ is the marginal likelihood serving as a normalisation term. Following the principle of Bayesian inference, the Bayesian brain hypothesis suggests that the brain is trying to infer hidden causes of sensory inputs by constructing a generative model which is modelling possible causes and the generative processes which describe how the hidden causes yield sensory data (likelihood). When new sensory evidence becomes available, the internal representation of hidden causes is updated using Bayesian inference. In this view, the human perceptual system can be seen as a "statistical inference engine" that extracts statistical and causal relations from the environment to infer hidden causes based on these learned relations (Clark, 2012).

The Bayesian brain hypothesis has been shown to be accurate in describing a wide range of human cognitive processes across different domains such as perception (Yuille & Kersten, 2006), motor control (Körding & Wolpert, 2004), memory (Anderson & Milson, 1989), and reasoning (Oaksford & Chater, 1994). For example, research has demonstrated its explanatory and predictive

power using sensory cue integration (Knill & Pouget, 2004). Several studies have found human subjects perform cue integration following Bayesian optimization (so called Bayesian ideal observer (Knill & Richards, 1996)) both within (Hillis, Watt, Landy, & Banks, 2004; Jacobs, 1999; Knill & Saunders, 2003) and across modalities (e.g. sight and touch or sight and sound) (Alais & Burr, 2004; Battaglia, Jacobs, & Aslin, 2003; Ernst & Banks, 2002; van Beers, Sittig, & van Der Gon, 1999). A study by Ernst and Banks (2002) is a classic example demonstrating how Bayesian inference accurately predicts behavioural performance.

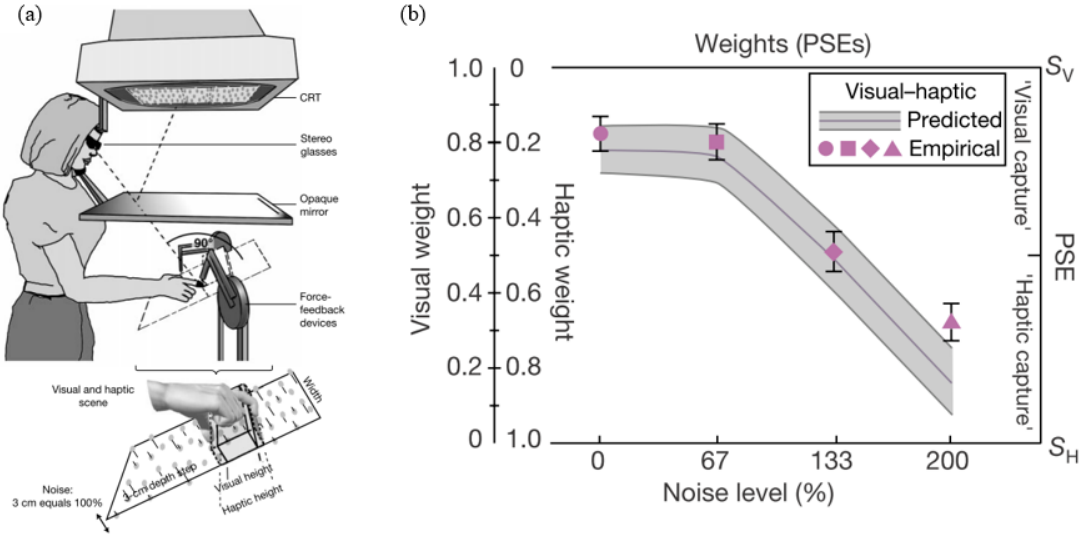

Figure 1-2 (a) The experimental setting and stimuli used in the study by Ernst and Banks, (2002). Participants were asked to integrate visual and haptic information to estimate the width of a virtual ridge. (b) Haptic and visual weights and Point of subjective equality (PSEs). The figure shows how the weighting between visual and haptic input changes with noise level for empirical and data predicted by Bayesian brain hypothesis. As the visual noise increases a stronger weighting is placed on haptic input. (figure adapted from Ernst & Banks, 2002).

In this study, human participants were required to judge which of two ridges was taller (Figure 1-2a). Three types of trials were presented. First, only haptic information was presented. Participants only received information by touching the ridge. Second, only visual information was presented. They were only provided visual information about the ridge. Third, they saw and

touched the ridge simultaneously and made integrated judgments based on both types of information. The Bayesian brain hypothesis suggests that, when multiple information sources become available to infer the hidden cause, the optimal method to integrate information is to compute a weighted average between estimations of signal information sources. The weights for information are determined by the relative reliabilities of the information sources. When one source is more reliable than another, the computation adds a stronger weighting to the more reliable source when computing the weighted average.

In Ernst and Banks' study, visual stimuli were presented with one of four different noise levels, i.e. manipulating the reliability of the visual cue. The results showed that behavioural performance in trials in which both visual and haptic cues were presented were very close to the predictions made by the Bayesian brain hypothesis (Figure 1-2b). This supports the Bayesian brain's assertion that multiple information sources provide the optimal method to integrate information. As can be seen, when the visual noise level increases, the weighting to haptic information also increases, again in line with the predictions of the Bayesian brain hypothesis. Similar results have also been found in other perceptual and sensory-motor tasks, e.g. colour consistency (Brainard & Freeman, 1997), motion perception, (Stocker & Simoncelli, 2006) and motion illusions (Weiss, Simoncelli, & Adelson, 2002). As a result of these and many other experiments the notion of perception as hypothesis testing as suggested by the Bayesian brain hypothesis has become a hot topic within Psychology and Neuroscience (Karl Friston, 2012; Gregory, 1968, 1980; Kersten, Mamassian, & Yuille, 2004).

1.1.2 Predictive coding: From Bayesian inference to Prediction

1.1.2.1 From Bayesian inference to predictive coding: A history of predictive coding

The notion of the brain as a hypothesis-testing machine is not a new idea, it has been around since the 60's and became a crucial concept in computational neuroscience (Lee & Mumford, 2003) and also in the Bayesian brain framework (Knill & Pouget, 2004). The Bayesian brain hypothesis is usually described at Marr's computational level (Marr, 1982). David Marr (1982) introduced a three-level framework to analyse a complex system:

- **Computational Level**: What is the goal of the computation, why it is appropriate, and what is the logic of the strategy by which it can be carried out?
- **Representation and Algorithm Level**: How can this computational theory be implemented? In particular, what is the representation of the input and output, and what is the algorithm for this transformation?
- **Hardware Implementation Level**: How can the representation and algorithm be physically realized?

The Bayesian brain hypothesis suggests a computational principle, i.e. Bayesian optimisation, as the ultimate goal of neural systems. Fundamentally, a computational description can be realised by multiple types of implementations. One particularly interesting formulation of the Bayesian Brain is predictive coding (Clark, 2012; Karl Friston, 2005; Hohwy, 2013; Seth, 2014). As mentioned above, the Bayesian brain hypothesis suggests that brains construct a generative model to infer the state of hidden causes. A generative model encodes the prior distribution of hidden

causes and the generative processes (i.e., the likelihood in Bayesian inference) which represent the compatibility of the sensory data with given hidden causes. This implies that generative models encode the joint probabilities between hidden causes and sensory data. Using these joint probabilities, the generative model can run internal simulations of the external generative processes. Therefore, a sensory system can generate predictions about sensory data by combining the current best estimation of the hidden causes with the joint probabilities. This allows a particular prediction-based form of Bayesian inference i.e. predictive coding. In this approach, generative models generate predictions about sensory input by simulating sensory data based on prior knowledge and generative processes. The predictions are then compared with actual sensory input. If the predicted sensory inputs do not fit the actual sensory inputs, a prediction error i.e. the difference between prediction and actual sensory evidence is computed. The model then updates its current estimation of the hidden sensory cause and generates new predictions about the sensory inputs in an attempt to minimise the current prediction error. By this process, statistical inference can be achieved and the model can find the best estimation of hidden causes through prediction error minimisation. Therefore, this predictive coding framework is identical to the view of "perception as a hypothesis testing process" (Gregory, 1980), in which perceptual systems test hypotheses by comparing predictions with actual sensory inputs. Predictive coding inherited all of the key characteristics of the Bayesian brain hypothesis. Importantly, predictive coding expanded the original framework by providing a physiologically plausible framework that was amenable to experimental investigation.

Recently the predictive coding framework became hugely influential in the modelling and description of human cognition and neural imaging data. Currently, many different computational predictive coding models have been proposed to interpret empirical neurophysiological data (Karl Friston, 2005; Gagnepain, Henson, & Davis, 2012; M.W. Spratling, 2008; Rao & Ballard, 1999;

Seth et al., 2012; Spratling, 2012; Wacongne, Changeux, & Dehaene, 2012). Next, I will briefly review some of the most influential predictive coding models.

Rao and Ballard's model

A milestone model proposed by Rao and Ballard attracted large amounts of attention in neuroscience. Rao and Ballard (1999) presented a predictive coding model of the visual cortex. The model is implemented by a hierarchical network and has a superior explanatory power for endstopping effect. In this model, higher level layers generate predictions of the responses of neurons in the next lower level (Figure 1-3a) and send predictions to lower layers through feedback connections. Lower level layers compute the residual (prediction error) between inhibitory signals, from higher level predictions (Figure 1-3b), and its current state. The residual signals are then sent back to the higher layers through feedforward connections. These residual signals are then used to correct the higher-level estimates of sensory inputs and make new predictions. In this model, a higher-level unit connects to multiple lower level units, and therefore, has a larger receptive field. Higher level units integrate signals from multiple lower level units to form a new prediction (Figure 1-3c).

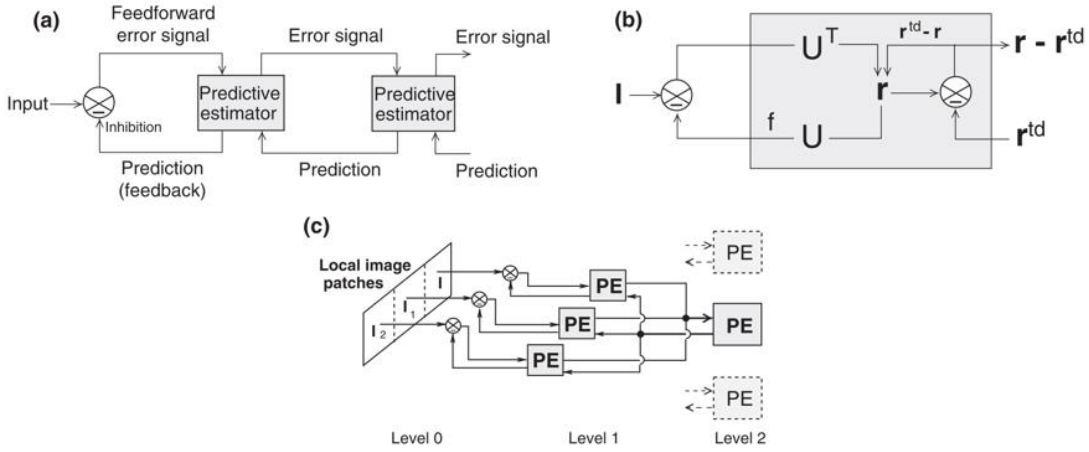

Figure 1-3 Hierarchical predictive coding model of the visual cortex by Rao and Ballard (1999) (a) architecture of the hierarchical predictive coding model of the visual cortex. Higher level predictive estimators (PE) predict the response value of the units in the next lower level. Lower level units compute the difference (prediction error) between actual activity and predictions from higher levels and send the prediction error signal to higher level units through feedforward pathways. (b) The structure of predictive estimators. I: actual sensory inputs. U and UT: synaptic weights; r: current estimate of the input signal. rtd: top-down predictions; (c) A three-level hierarchical network. Three image patches at level 0 are processed by three level 1 PE units. Three level 1 PEs connect to one level 2 PE, suggesting a larger receptive field for higher level units (figure adapted from Huang & Rao, 2011)

This architecture suggests that the top-down signal is crucial for the computation of perceptual information and determines the state of lower layers by the predictions at high level layers through feedback connections. Therefore, it can naturally explain several characteristics of V1 neurons, for example, the so-called end stopping effect. The end stopping effect was first described by Hubel and Wiesel (Hubel & Wiesel, 1965). Some neurons (end-stopping neurons) in V1 activate more when a contour is within its receptive field. However, their responses become smaller when the contour exceeds its receptive field, suggesting suppression from other neurons with receptive fields close to this neuron. This effect is well replicated by Rao and Ballard's model. When the contour is longer than the receptive field, higher level layers can successfully predict the content of this receptive field by receiving information from neurons surrounding this receptive field. When the prediction from higher level layers mirrors the content of the input of this neuron it generates a smaller prediction error yielding a smaller activation (Figure 1-4a). Conversely, when the contour only falls within the receptive field of a single neuron, the high level layer has no information to predict the "atypical" input. Therefore, the higher level layer fails to predict the input content. This drives the low level neuron to produce a large prediction error and therefore shows a larger activation (Figure 1-4b). In summary, this model parsimoniously explains the endstopping effect within a predictive processing framework.

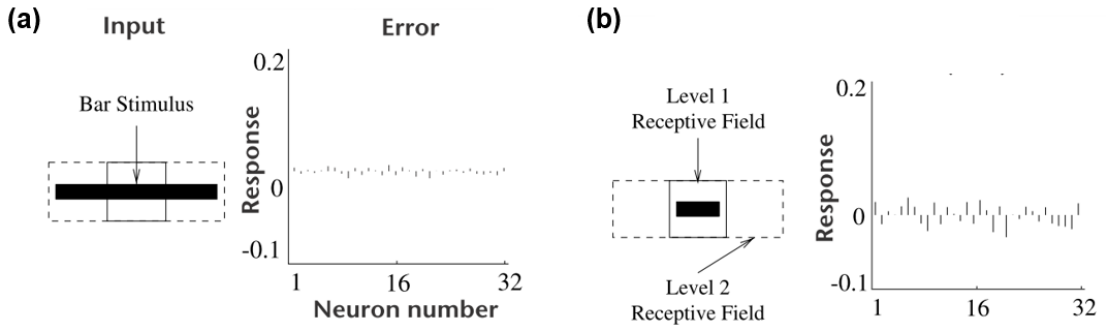

Figure 1-4 The end stopping effect simulated with the predictive coding model for visual cortex (Rao & Ballard, 1999) (a) a bar stimulus exceeds the central receptive field. Due to the contents of the central receptive field being predicted from the content of the neighbouring receptive fields, the prediction error generated from the central neuron is small. (b) a bar stimulus only presented in the central receptive field. Due to no predictive visual information from neighbouring receptive field, the prediction error is large for the central neuron. (figure adapted from Rao & Ballard, 1999)

Karl Friston's model of predictive processing

The core idea underlying the Bayesian brain and predictive coding frameworks is that the brain is an inference engine trying to optimize probabilistic representations of what caused its sensory input. Another generalisation form of predictive processing was proposed by Karl Friston at University College London (Karl Friston, 2005, 2010). His addition to the predictive processing framework was to add a more fundamental principle called the "free energy principle" that links predictive processing to statistical physics and information theory. Free energy can be seen as a measure that limits surprise when receiving new data into a generative model. Surprise is simply the improbability of sensory data, given a model of the environment. Therefore, prediction error minimization in predictive processing is equivalent to free energy minimization. Statistical inference and learning of a biological system can be achieved by minimizing the free energy of the internal generative models. Free-energy principle entails the Bayesian brain hypothesis and

predictive processing and can be seen as a biologically plausible implementation of these frameworks.

This section has highlighted two key models that have been influential within the predictive processing field, following these models many other predictive processing models have been proposed that exhibit strong explanatory power across a wide range of domains (Baldeweg, 2007; Seth, 2014; Seth et al., 2012; Wacongne et al., 2011).

Despite the predictive processing framework being applied to a wide range of behavioural and neurophysiological data, and also to phenomenological properties of consciousness such as emotion, feelings of presence, and the construction of the self, the role of predictive processing in conscious perception remains unclear. The next section will move on to review literature concerning the relationship between predictive processing and conscious content.

1.2 Predictive processing and Conscious Content

Broadly, consciousness can be broken down into two major components: arousal, which ranges from unconsciousness to full waking consciousness, and conscious content, the current content of a person's experience This thesis focuses on conscious content within the predictive processing framework.

Even early formulations of the Bayesian brain hypothesis describe conscious experience as the result of Bayesian inference. According to this view information processing and computation for perception are *unconscious inference* (Gregory, 1980; Helmholtz, 1866; MacKay, 1956; Neisser, 1967; Rock, 1983). Only the results (best estimation of hidden causes) of statistical inference reach conscious awareness and form the content of conscious experience. The factors that render the computational process under statistical inference unconscious are still unknown.

The motion illusion in aperture problem provides a good example of unconscious inference (Horn & Schunck, 1981; Wuerger, Shapley, & Rubin, 1996). The aperture problem is confronted when using local information to determine a global pattern. Figure 1-5 shows a grating pattern presented in an "aperture". When the grating is moving, a certain moving direction can be perceived (usually perpendicular to the line direction of the grating). However, the motion information is ambiguous locally because the motion pattern can be caused by infinite possibilities of global motion. For example, if the lines of the grating are moving to the left in the aperture, global patterns which move left, up-left, and up can produce the same spatial and temporal visual inputs (Figure 1-5b). Therefore, it is impossible to determine the global motion direction by local information alone.

However, even though the local information is ambiguous, observers feel no ambiguity about the direction of movement, they instantly experience a certain motion direction. This observation implies that the computations underlying the orientation of movement is fully unconscious, therefore unconscious inference must play a crucial role in what we perceive consciously.

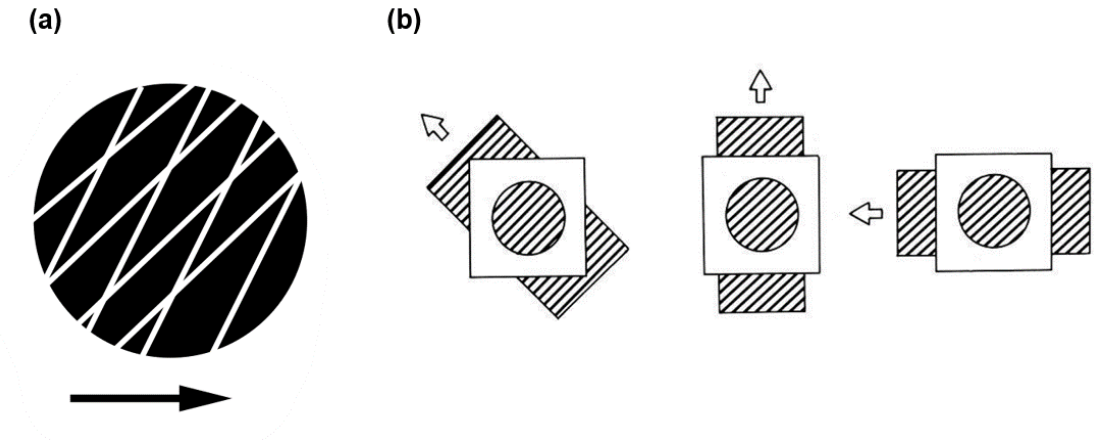

Figure 1-5 The motion illusion in aperture problem. (a) shows a grating pattern presented in an "aperture". The grating pattern is moving globally, but observers can only watch the pattern through the aperture. (b) A moving pattern in the aperture can be generated by infinite global moving directions, for instance the three examples in this figure. Observers usually perceive the

moving direction is perpendicular to the grating line directions. (figure adapted from Weiss et al., 2002).

To explain how the visual system solves the aperture problem, Weiss, Simoncelli, and Adelson proposed a model using Bayesian inference that describes conscious percepts of the motion direction in the aperture problem (Weiss et al., 2002). Following the Bayesian inference, the model multiplies prior of motion speed (a probability distribution that encodes previous experience) and a likelihood function that describes the probability of motion speed given the local motion pattern in the aperture. The model assumes a simple Gaussian prior which favours slow motion, i.e. higher probabilities for low velocities than high velocities. The likelihood incorporates sensory precision (inverse of sensory noise level). The model predicts that when the image contrast is low the precision should also be low and yield less weighting on likelihood. On the other hand, when the image contrast is high, the sensory signal is reliable, thus yielding higher weighing on likelihood. The final perceptual experience is determined by maximizing the posterior, the hypothesis with highest probability in the posterior distribution. This simple Bayesian model beautifully captures how the conscious experience of motion direction is affected by contrast, edge orientation, and other stimulus features, including the classic observation that stimuli appear to move slower at lower contrasts and vice-versa (Figure 1-6). The model also predicts a wide range of unintuitive motion phenomena, which are usually attributed to different neural mechanisms.

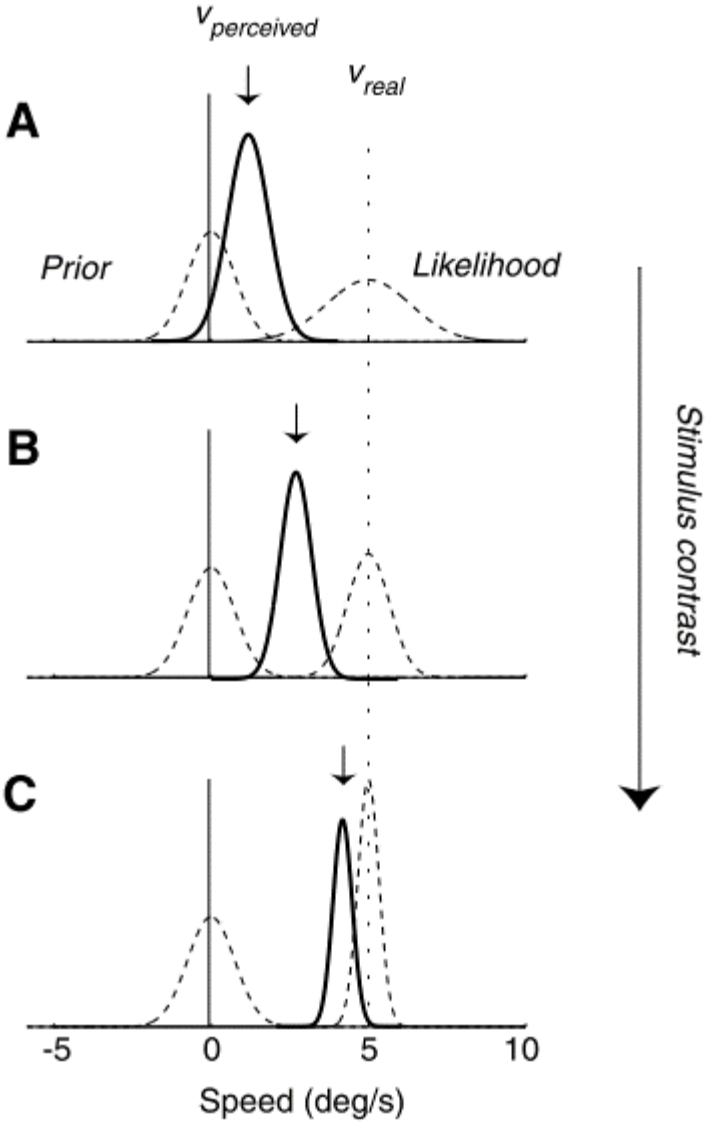

Figure 1-6 Bayesian inference relating to motion speed in the aperture problem. Stimulus contrast in an aperture determines the precision of likelihood. As a result, the perceived speed using Bayesian inference is modulated by stimulus contrast. A, B, and C illustrate perceived speed using three different stimulus contrast values (figure adapted from Hürlimann, Kiper, & Carandini, 2002).

In this example, conscious experience of motion is the final decision of Bayesian inference. More importantly, the complicated neural processes underlying the computation are utterly

unconscious. At the conscious level, we are unaware of how the brain integrates multiple features using Bayesian rules and computes the posterior distribution of motion directions. We also cannot intentionally and consciously access the prior distribution, which should be represented as long-term knowledge in the neural system. Therefore, a possible approach to understanding the neural basis of conscious perception is to investigate how unconscious inference is able to influence conscious perceptual decisions.

The notion of unconscious perceptual decision making has been addressed throughout the history of visual science since e.g. *Optics of Ptolemy 100-170 A.D.* (Smith, 1996). Nevertheless, there is still no theory providing a fully explanatory account of how the result of the unconscious inference generates phenomenal experience.

As Hatfield states (2005, p. 120): *"to be fully explanatory, unconscious inference theories of perception must explain how the conclusion of an inference about size and distance leads to the experience of an object as having a certain size and being at a certain distance, or how a conclusion about a spectral reflectance distribution yields the experience of a specific hue. In other words, the theories need to explain how the conclusion to an inference, perhaps conceived linguistically, can be or can cause visual experience, with its imagistic quality. This is the phenomenal experience problem"*

As the explanatory power of predictive processing continues to become more prominent within neuroscientific and consciousness research, it is increasingly essential to understand this relationship between unconscious statistical inference and conscious experience.

This section describes how predictive processing plays a crucial role in perception through unconscious perceptual inference. The next section will introduce and review the hierarchical predictive processing framework, which states that the same principle of perceptual inference can be used across sensory hierarchies in order to process differing levels of information.

1.3 Hierarchical predictive processing

Predictive processing has been proposed as a unifying principle underlying all neural processing, though whether this ambitious claim stands up in practice remains to be seen. The human visual system, for example, is organized hierarchically and has extensive reciprocal cortico-cortical connections between different cortical levels (Felleman & Essen, 1991). Predictive processing models of vision therefore needed to address the hierarchical structure of neural systems and as a result have been primarily discussed in terms of a hierarchical structure (Clark, 2012; Karl Friston, 2002; Huang & Rao, 2011; Mumford, 1992). For example, the predictive processing model of the visual cortex proposed by Rao and Ballard, is modelled as a hierarchical network, with higher level units attempting to predict the responses of units in the next lower level. In the hierarchical predictive processing framework, prediction units at higher levels within the model aim to predict the states of units at the next lower level through a generative process. Prediction errors at lower levels within the model are computed by comparing the prediction generated by the higher level and the actual states of the lower level units. The lower level units pass the prediction error signal through feedforward pathways to higher level units, which has the effect of adjusting estimations of hidden causes and modifying predictions. At the lowest level of this hierarchy, the prediction units are tasked with predicting actual sensory inputs. If these predictions do not fit the sensory inputs, the prediction errors propagate from lower levels to higher levels one by one through feedforward connections between each level. This process iteratively occurs until prediction errors have been minimised across the whole hierarchy, which has the effect that the prediction closely mirrors the actual input, thus completing the inference process.

Deep hierarchical generative models outperform models with single or few levels (Hinton, 2007). Due to this multilevel structure, hierarchical predictive processing is capable of modelling the generative process of external multilevel causal structures and is able to infer abstract hidden causes in the environment (Karl Friston, 2005; Hohwy, 2014). The hierarchical structure progressively encodes more complex features from lower levels to higher levels and thus is able to learn complicated, non-linear relationships in real world scenarios. This hierarchical structure also implies hidden causes at different spatial and temporal scales can be inferred and predicted. In terms of spatial scale, lower level units predict a smaller spatial input area, whereas higher level units encode statistical relationships between several lower level units. Therefore, predictions from higher level units cover a larger input area compared to lower level units. Rao and Ballard (1999), demonstrated that receptive fields of higher level units are larger than lower level units in a hierarchical predictive coding model (Figure 1-3c and Figure 1-7).

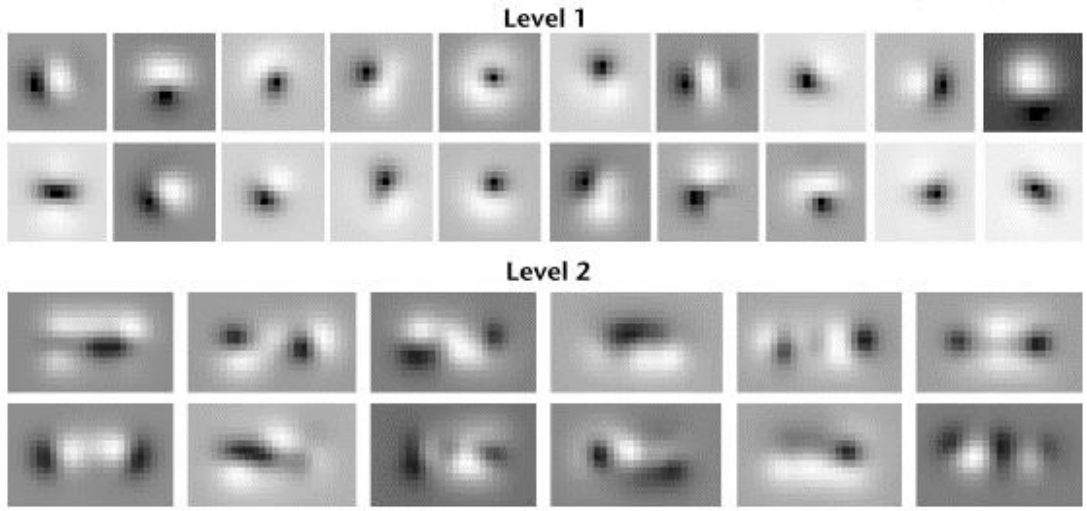

Figure 1-7 Learned synaptic weights level-1 and level-2 modules in the hierarchical predictive coding model of the visual cortex. The receptive fields of neurons at level-2 are larger than the receptive fields of level-1 neurons, consistent with human and animal visual processing hierarchy. (taken from Rao & Ballard, 1999, Fig2).

In terms of temporal scales, lower levels learn and predict fast changing regularities such as contour and orientation. While higher levels generate predictions about longer-term invariant regularities such as faces (Hohwy, 2014). An example by Friston et. al., (2009) showed that information of size and strength of a bird can be extracted from bird songs, indicating that hidden causes in a slower time scale (bird identity) can be inferred by a fast changing time scale (bird songs).

Wacongne et al. directly tested the hierarchical predictive coding hypothesis using a hierarchical auditory novelty paradigm using EEG and magnetoencephalography (MEG) (Wacongne et al., 2011). Participants were presented with blocks of auditory sound strings consisting of a rare sequence pattern (xxxxX) and a frequent pattern (xxxxY). This design creates local and global regularities. The local regularity is established by the standard tones (x) which made up the majority of the string. Therefore, when a deviant tone (Y) was presented, it violated the local regularity and caused a local prediction error. The global regularity was established by presenting a frequent pattern (xxxxY) in a block. Presenting a rare sequence pattern (xxxxX) violated the global regularity and caused a global prediction error. Wacongne et al. examined the human neural correlates of local and global prediction errors using EEG and MEG. The results show that the local deviant tone (Y) elicited a Mismatch negativity (MMN), a neurophysiological index of auditory change detection, even though the standard pattern (xxxxY) was globally predictable, suggesting that the MMN was sensitive to the local regularity but insensitive to the global regularity pattern. A temporally late mismatch response was found only responding to the violation of the global pattern (xxxxX) even though the stimulus was locally standard (the fifth stimulus X in xxxxX). This suggests a higher-order regularity learning mechanism exists which generates high-level predictions according to global regularities. To further examine hierarchical predictive coding, Wacongne and her colleagues examined the neural responses elicited by

stimulus omission. As mentioned above, the hierarchical predictive coding framework suggests that higher level predictive units generate predictions about the state of units in the next lower level. Therefore, high level predictions indirectly influence low level predictions via feedback connections down to the lowest level of the sensory system. To test this hypothesis, the authors designed a stimulus omission condition in which an expected stimulus was omitted to test top-down prediction. The omissions of the fifth stimulus in xxxxY block was compared with the one in xxxxX block. In the xxxxX, the omission of the fifth stimulus (X, the local standard) violates only local regularities. In contrast, in xxxxY block, the local deviant stimulus (Y) was expected to occur through high level prediction after four local standard stimuli (x). Therefore, the omission of the local deviant stimulus (Y) violates both local and global predictions. The result showed that the omission in the xxxxY block elicited a larger temporally early mismatch response than the omission in the xxxxX block, supporting the prediction by hierarchical predictive coding framework.

To summarize, hierarchical predictive coding is capable of encoding a wide range of information types from concrete to abstract, using low to high levels within the hierarchy. The benefits of a hierarchical predictive coding model have been demonstrated through numerous behavioural and neuroimaging studies (Tenenbaum, Kemp, Griffiths, & Goodman, 2011; Tervo, Tenenbaum, & Gershman, 2016; Wacongne et al., 2011). In terms of information processing within a sensory hierarchy questions remain about the effects that differing levels of sensory information i.e. low, cross-modal and abstract levels of information, have on predictive processing. An important feature of these models is that they have been shown to learn regularities and predict hidden causes at different spatial and temporal scales. Next I will review the evidence showing predictive processing is involved in different types of regularity learning, highlighting an often overlooked point, that in order to make a prediction about the future, first a system must learn

environmental regularities.

1.4 Predictive processing and Regularity Learning

The operation of predictive processing relies largely on adequate knowledge of environmental regularities. Effective predictions can only be made when the environmental regularities are successfully extracted and learned (Seriès & Seitz, 2013). Theorists have suggested that a system using predictive processing should be able to learn environmental regularities and model the causal regularities of the external world (Hohwy, 2014). Empirically, previous research has showed that the human brain is capable of learning complex statistical relationships (e.g. statistical learning, Fiser, Berkes, Orbán, & Lengyel, 2010). However, it is still unclear to what extent the regularities in the environment can be learned and incorporated into predictive processing to predict future events. In this thesis, I focus on three types of regularity learning: associative learning (Chapter 2), sequence learning (Chapter 3), and stimulus repetition (Chapter 4). The following are brief introductions to previous research into the relationship between predictive processing and these three types of regularity learning.

1.4.1 Associative learning

There is a growing body of evidence supporting the claim that neural responses of associative learning are compatible with the predictive processing framework (Kok, Brouwer, Gerven, & de Lange, 2013; von Kriegstein & Giraud, 2006). A good example carried out by Kok, Jehee, and

de Lange demonstrated that during an auditory-visual cross-modal association task there was task relevant predictive information located within the visual cortex (Kok, Jehee, & de Lange, 2012) using fMRI. In this study participants were required to compare orientations or contrast (in different blocks) of two consecutive grating stimuli in each trial. The two grating stimuli were always preceded by an auditory tone (the cue). The auditory cue consisted of either a low or high tone which 75% of the time predicted the orientation of the subsequent grating stimuli. Therefore, after several exposures, participants learned this audio-visual association. With multivariate pattern analysis (MVPA), they found when the orientations of the grating stimulus were predicted by the cue, MVPA classifiers were better able to classify the orientation from the visual cortex compared to unpredicted orientations, suggesting that the cross-modal association shaped the representation within the visual cortex. This result is consistent with the expectations of the predictive processing framework, which posits that high-level cortical areas can predict and modulate the states of low-level neurons. More importantly, Kok et al. found that the amplitude of BOLD signals was higher when orientations of the grating stimulus were unexpected compared to expected. A finding that is again compatible with a predictive processing framework and suggests the neural processing underlying the BOLD signals were driven by the prediction error. When the prediction is not consistent with the visual input, low-level visual areas generate larger prediction error signals. Additionally, a study by den Ouden et al. showed similar “prediction suppression” and prediction-error related neural activation in a cross-modal associative learning task (den Ouden, Friston, Daw, McIntosh, & Stephan, 2009). Activity in primary visual cortex and putamen progressively positive correlated learning-dependent surprise. Together, the results of these studies provide support for the predictive processing framework and suggest that at least some forms of associative learning may be driven by predictive processing.

1.4.2 Sequence learning

Sequence learning is a well-studied ability among human and other animals (Conway & Christiansen, 2001; Goldstein et al., 2010). Neural systems are capable of learning consecutive events with deterministic or probabilistic spatial and temporal relationships. From a predictive processing point of view sequence learning is thought to occur due to a generative model, which encodes the transitional probabilities between consecutive events from the sequence history. Therefore, based on sequence histories, predictive processing is able to generate predictions about incoming sensory events. Recent evidence shows that when a sensory event is predictable, due to regularities within the sequence pattern, the neural responses elicited by an expected event is suppressed, a finding expected within the predictive processing framework (Karl Friston, 2005). For example, a study by Aizenstein et al., shows that implicit sequence expectation about colour and location is able to supress neural responses to visual input. In this study participants viewed consecutive sequences consisting of a series of items with two attributes, colour and location (Aizenstein et al., 2004). They were required to press different buttons according to the colour or shape of an item. The colours and shapes of items were determined independently and follow two different first-order Markov chains (see Figure 1-8). Therefore, this design allows for two sequence regularities to be embedded in the sequence, task-relevant (explicit) and task-irrelevant (implicit) features. They also monitored the awareness of the sequence knowledge for each participants during the whole experiment. The results showed that when the task-irrelevant feature was consistent with prediction visual regions (V1, V2, and V3) showed decreased activity, suggesting the implicit sequence learning may involve predictive processing.

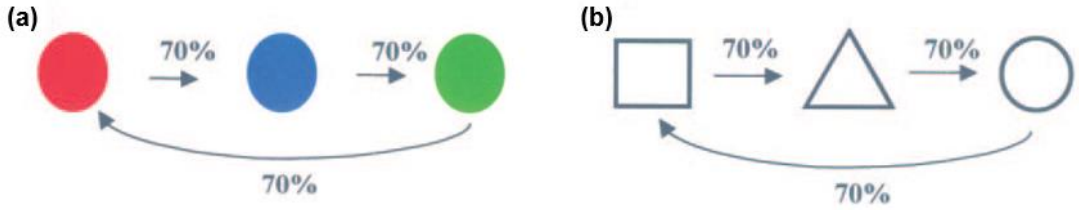

Figure 1-8 The first-order Markov chain transition rules used by Aizenstein et al. (2004). Objects in a sequences have two changing attributes, colour and location. The colours and shapes of items were determined independently and follow two different first-order Markov chains (a and b). (figure adapted from Aizenstein et al., 2004)

Similarly, Huettel, Song, and McCarthy manipulated sequence uncertainty of a binary sequence of eight stimuli so that the uncertainty changed dynamically over time (Huettel, Song, & McCarthy, 2005). They found activity in prefrontal, parietal, and insular cortices increased with increasing uncertainty, suggesting that predictable sequences elicited lower neural responses.

A more recent study by Davis and Hasson (Davis & Hasson, 2016) showed that not only low-level sensory features, such as colour, shape, etc., but also high-level expected semantic features caused a reduction in activation. Suggesting that a marker of predictive processing can also be found when a sequence has high-level semantic regularities. The transition probabilities were determined by two first-order Markov chains for the location and semantic category of the next image. There were three conditions in this study. The sequence was embedded with either the location, semantic, or both regularities. All three conditions showed reduction of the BOLD response in left rostral anterior cingulate cortex, bilateral putamen, caudate, thalamus, right precentral gyrus, and left visual cortex. More importantly, the predictability of the semantic category decreased BOLD activity in areas of the ventral visual stream and lateral temporal cortex which is considered to be a key semantic processing area. Predictability of location of stimuli decreased activation in dorsal fronto-parietal areas which are thought to be responsible for the endogenous control of spatial attention.

In another study by Reithler, van Mier, and Goebel, participants learned movement sequences by continuously tracing paths of different predefined mazes (Reithler, van Mier, & Goebel, 2010). They found BOLD signals of learning-related networks of cortical areas decreased over time during motor sequence learning. Further analyses showed decreased activation when task performance increased in bilateral supplementary motor areas (SMA), posterior intraparietal sulcus, and anterior insula. The authors state that continuous motor sequence learning yields efficient neural processing in recruited cortical areas. The result can be also interpreted by a reduction in prediction errors that are generated when movements more precisely follow the maze paths. A similar result was found by Steele and Penhune (2010). They found that across several days of motor sequence learning, global BOLD signals decreased along with behavioural performance improved in the network for motor learning including cerebellum, premotor cortex, basal ganglia, pre-supplementary motor area, and supplementary motor area (Steele & Penhune, 2010).

1.4.3 Stimulus repetition

Stimulus repetition is the simplest form of statistical regularity in the environment. Previous research has shown neural systems are sensitive to this form of regularity (Kalanit Grill-Spector, Henson, & Martin, 2006). Here, I review two well studied neural correlates of stimulus repetition, repetition suppression and the odd-ball effect.

1.4.3.1 Repetition suppression

Repetition suppression refers to a reduction in neural activation when stimuli are repeated (Buckner et al., 1998; K. Grill-Spector et al., 1999; Henson, Shallice, & Dolan, 2000; Kourtzi & Kanwisher, 2001; Naccache & Dehaene, 2001; van Turennout, Ellmore, & Martin, 2000; Wiggs & Martin, 1998). Studies have demonstrated repetition suppression for different representations

across levels of the visual processing hierarchy (Kalanit Grill-Spector & Malach, 2001), suggesting a ubiquitous characteristic in neural systems. An earlier theory of repetition suppression suggests that decreases in activation are due to neuron fatigue as a result of stimulus repetitions (Kalanit Grill-Spector et al., 2006) . This point of view only considers bottom-up information flow through sensory cortex.

In contrast to these early theories, Summerfield and his colleagues showed that expectation plays a role in repetition suppression (Summerfield et al., 2008). In this study, the probability of stimulus (face) repetition was manipulated. Participants were required to identity inverted faces in a stimulus stream. In one block, the probability of face repetition in two consecutive presentations was high (75%). Therefore, after exposures to several presentations, participants developed an expectation about stimulus repetition. In the other block, stimulus repetition rate was only 25%, and hence, repetitions were not expected. The result showed that repetition suppression in the fusiform face area (FFA) was significantly larger when stimulus repetitions were expected compared to unexpected. This finding is consistent with the predictive processing framework suggesting top-down predictions are crucial in modulating neural responses for sensory inputs. When repetition is expected, the visual system is able to generate more precise predictions about the upcoming face stimulus, resulting in less prediction error. In contrast with the neuron fatigue theory of repetition suppression, the predictive processing framework more parsimoniously explains how expectation influences neural responses during repetition of sensory information. In sum, the study of repetition suppression supports the notion that statistical regularities can be extracted from the environment by the neural systems yielding more precise predictions and lower prediction errors.

1.4.3.2 The odd-ball effect

In contrast to repetition suppression, the neural correlates of an unexpected sensory event are usually tested with an odd-ball paradigm (Squires, Squires, & Hillyard, 1975). In the odd-ball paradigm, successive presentations of frequent stimuli (standard stimuli) are interrupted by infrequent stimuli (deviant stimuli). The infrequent stimuli elicit a "surprise response" in neural systems. By measuring the surprise responses, one can examine neural mechanisms of expectation and novelty (e.g., Folstein & Van Petten, 2008; Hruby & Marsalek, 2002).

To interpret the result of the odd-ball paradigm within a predictive processing framework, when a deviant stimulus is detected, the change of stimulus property causes a prediction error which elicits a larger neural response than elicited by standard stimuli.

One notable example of the odd-ball effect within the auditory domain is the mismatch negativity (MMN). MMN is an event-related potential (ERP) component usually shows in the auditory odd-ball paradigm. When human subjects are presented with a series of auditory stimuli consisting of standard and deviant stimuli, the deviant stimuli elicit larger electric potentials than the standard stimuli. The peak of MMN occurs about 100–200 ms from the onset of deviant stimuli (Garrido, Kilner, Stephan, & Friston, 2009; Näätänen, 1995). MMN can be elicited without attention (Näätänen, Paavilainen, Rinne, & Alho, 2007; Näätänen, Paavilainen, Titinen, Jiang, & Alho, 1993, but still under debate, see Sussman, 2007) and can even be found in coma patients (Fischer et al., 1999; Fischer, Morlet, & Giard, 2000; Morlet, Bouchet, & Fischer, 2000; Näätänen & Escera, 2000; Wijnen, van Boxtel, Eilander, & de Gelder, 2007). Therefore, MMN has been considered as a pre-attentive neural response.

Similar to auditory stimulation, growing evidence shows visual odd-ball stimuli are also able to induce mismatch ERPs. The finding of visual mismatch negativity (vMMN) shows similar

novelty detection mechanisms between auditory and visual information processing (Stefanics, Kremláček, & Czigler, 2014).

Early theories of MMN focus on two interpretations. The *model-adjustment hypothesis* suggests that MMN was driven by neural processes of the auditory change detection process, located in temporal and frontal regions (Näätänen, Jacobsen, & Winkler, 2005; Tiitinen, May, Reinikainen, & Näätänen, 1994). This theory emphasises a pre-attentive sensory memory mechanism that learns auditory regularities and auditory context in the environment. MMN then is driven by a comparison computation in which the current auditory input is compared with the memory trace of previous inputs (Näätänen, 1992). According this hypothesis, the MMN is thought to reflect a perceptual model updating processes.

Another theory of the MMN suggests that it is driven by the adaptation of local neurons in the auditory cortex. The *adaptation hypothesis* is proposed by Jääskeläinen et al (2004), who suggests that adaptation to repeated auditory input causes attenuation and delay of the N1 (Loveless, Levänen, Jousmäki, Sams, & Hari, 1996). In contrast to the model-adjustment hypothesis which emphasises novelty detection, the adaptation hypothesis argues that MMN is caused by adaptation to standard stimuli. The N1 response to frequent auditory input is delayed and suppressed with stimulus repetition. Horváth et al., provides evidence that the amplitude of N1 component is affected by deviant stimuli (Horváth et al., 2008). Therefore, suggesting that the MMN may be due to adaptation of local neurons although the adaptation hypothesis is still not widely accepted (see Näätänen et al., 2005).

Predictive processing has been considered as a candidate mechanism of MMN. According to predictive processing, the auditory system tracks input regularities and makes predictions. A comparison between bottom-up inputs and top-down predictions result in computation of prediction errors. The resulting comparison (prediction error) is then used to update internal models

encoding input regularities. In this framework, MMN is thought to be driven by neural response of prediction error. During the repetitions, the sensory inputs can be better predicted. This results in lower prediction error and absence of MMN (Baldeweg, 2006; Karl Friston, 2005).

The predictive processing framework shares several properties with the model-adjustment hypothesis. Both theories describe an internal model which learns regularities and makes predictions about sensory inputs. Predictive processing further provides a neuronally plausible and computational framework for MMN. Friston (2005) therefore considers MMN as neural evidence of predictive processing in his model.

Garrido et al. (2009) re-explained possible MMN mechanisms and proposed that predictive processing can also explain the contribution from neural adaptation to MMN as described by the adaptation hypothesis. The Bayesian inference nature of predictive processing considers precision of information source in inference processes. The result of inference is a weighted average between top-down prior and bottom-up sensory input based on their relative precisions. Therefore, when standard stimuli are repeated, top-down prior can more precisely predict sensory inputs. This results in more weight being assigned to top-down signals and less weight assigned to bottom-up input, causing less impact from sensory inputs. This prediction from predictive processing is consistent with the adaptation hypothesis. Therefore, predictive processing provides an integrated view of MMN that unifies the model-adjustment and the adaptation hypotheses. Garrido et al. directly tested potential computational models of MMN by Bayesian model comparison (Garrido et al., 2008). The result showing that, as predictive processing expected, the combination of the model-adjustment hypothesis and the adaptation hypothesis best explained the generation of MMN. More recently, as mentioned in section 2.3, Wacongne et al (2011) provided direct evidence showing that the MMN amplitudes in a hierarchical odd-ball paradigm can be best explained by the hierarchical predictive coding framework.

In summary, the predictive coding framework is able to combine apparently distinct models of the MMN, which embodies both adaptation and model-adjustment, while also reconciling previous empirical findings relating to MMN.

1.5 Motivation and outline of dissertation

As can be seen from the literature review, over the last two decades there has been considerable advance in theoretical frameworks and accumulating empirical evidence supporting the predictive processing framework. These breakthroughs have highlighted the explanatory power of this framework when describing neurophysiological responses, and even giving insights into the anatomical organization of the human brain.

However, crucially we still have a limited understanding about fundamental aspects of predictive processing, such as how it deals with different types of information, learns statistical regularities and perhaps most fundamentally of all what its relationship to conscious experience is. The predictive processing framework has even been touted as a candidate unified theory of the brain (Karl Friston, 2010), however it still requires a systematic investigation into the role that predictive processing plays across different brain functions for this claim to be accepted. While validating a predictive processing based unified theory of the brain is beyond the scope of a single thesis, this thesis aims to investigate the major gaps in our current understanding of the predictive processing framework via three broad research topics:

1.The relationship between unconscious statistical inference reflected by predictive processing and conscious access.

2. The relationship between differing levels of information i.e. low level, cross-modal, etc. within the sensory hierarchy and the predictive processing framework.

3. Differing forms of regularity learning used to make predictions about future events within the predictive processing framework.

The three topics will be addressed across three studies making up Chapters 2, 3, and 4 of this thesis. Briefly, I will now describe the motivation behind each chapter.

1.5.1 Chapter 2 (study 1): Predictive processing and conscious access

In chapter 2, I investigate the relationship between predictive processing and conscious content. I examine whether predictions about visual events caused by auditory cues could facilitate conscious access to a visual stimulus. I trained participants to learn cross-modal associations between auditory cues and colour changes of a visual stimulus. I hypothesised that conscious access to such representations could be influenced by cross-modal cues if strong associations were formed via extensive training. To test this, I delivered an auditory cue while a visual target was rendered subjectively invisible via MIB, and I then gradually changed the colour of the target to assess how the congruency between the cue and the visual sensory event influenced the timing of subjective target reappearance. Timing was compared with that from control trials in which the target was physically removed. In this way, it was possible to quantify the effects of prediction on awareness while minimizing the impact of other cognitive processes such as attention and response bias.

1.5.2 Chapter 3 (study 2): Predictive processing with low-level visual information

In Chapter 3, I investigated how low level information within the sensory hierarchy is dealt with by the predictive processing and its relationship to regularity learning, and how this regularity learning can be used to make predictions about future events. A sign that the visual system actively processes and predicts sensory signals is revealed by the recently discovered ‘perceptual echo’, in

which the cross-correlation between a sequence of randomly fluctuating luminance values and occipital electrophysiological (EEG) signals exhibits a long-lasting periodic (~100ms cycle) reverberation of the input stimulus (VanRullen & Macdonald, 2012). As yet, the mechanisms underlying the perceptual echo and its function are unknown. Reasoning that natural visual signals often contain temporally predictable, non-periodic features, we hypothesised that the perceptual echo may reflect an underlying iterative learning process (governed by predictive processing) that is associated with regularity learning. To test this idea, I presented subjects with successive repetitions of a rapid non-periodic luminance sequences, and examined the modulation from sequence repetition to perceptual echo. To investigate the perceptual echo further I created a predictive processing model to simulate and explain key features of the perceptual echo in order to validate that the empirical findings of this could indeed be accounted for by predictive processing mechanisms.

1.5.3 Chapter 4 (study 3): Temporal predictive processing using supra-modal information

In chapter 4, I investigated how differing levels of information, this time regularity learning of temporal duration is dealt with by predictive processing and how this form of regularity learning is used to make predictions about future events. Temporal duration comprises abstract and supra-modal information that is independent from modality-dependent sensory features. Therefore, temporal duration allows us to understand how predictive processing works on high-level, abstract information. In this study, I examined predictive processing using a duration odd-ball paradigm (Chen, Huang, Luo, Peng, & Liu, 2010; Jacobsen & Schröger, 2003; Näätänen, Paavilainen, & Reinikainen, 1989; Tse & Penney, 2006; see also review Ng & Penney, 2014) and examined neural responses using EEG. I used a re-designed version of the duration odd-ball paradigm to isolate the neural correlates of predictive processing for duration information. I then used event-related

potentials (ERPs) and state-of-the-art multivariate pattern analyses (MVPA) to investigate the supra-modal nature of temporal predictive processing. Finally, I investigated the neural correlates of a high-level temporal regularity learning and prediction mechanisms that receives temporal information from visual and auditory inputs and integrates duration and rhythmic information to determine a unified temporal prediction about future events.

1.5.4 An integrated approach

The three studies were designed to specifically address how predictive processing deals with diverse types of predictive information in different contexts and tasks to further provide a broader understanding of the predictive processing framework.

1.5.4.1 Conscious access

To investigate the relationship between predictive processing and conscious content, Chapter 2 focuses on quantifying the effects of prediction on conscious access, while Chapter 3 examines how low-level visual sequence learning that occurs unconsciously affects predictive processing. Chapter 4, investigates the temporal domain using information that is both abstract and supra-modal, using a paradigm that has been shown to automatically detect change in sensory features without the need for conscious access. Therefore, for each of these experiments it is possible to examine the mechanisms of predictive processing in operation at differing conscious levels.

1.5.4.2 Sensory hierarchy

In terms of information processing in a sensory hierarchy, Chapter 2 focused on cross-modal predictive information. Chapter 3 then examined low-level visual luminance dynamics, predictive information was embedded in the repetition of the modality-specific low-level visual features. In

Chapter 4 the predictive information was supra-modal duration information, in this case the neural system needs to extract the abstract temporal information to learn the regularity of duration repetition. This thesis will examine predictive processing through the three different levels of sensory information (modality-specific, cross-modal, and supra-modal).

1.5.4.3 Regularity learning

In terms of how different formats of regularity learning affect predictive processing, in the thesis I examined predictive processing via three common types of regulatory leaning utilized in previous studies from psychology and neuroscience. Chapter 2 focuses on predictive processing using associative learning. Chapter 3 examined predictive processing in sequence learning and sequence repetition. In Chapter 4, I examined predictive processing by stimulus repetition. The three types of regularity learning cover a broad range of cognitive aspects used in learning predictive information. I will investigate and discuss how the three types of regularity learning contribute to predictive processing.

1.6 Unique Contribution

Overall this thesis provides a global picture of predictive processing that aims to understand how the brain uses different types of predictive information in different contexts and tasks. This thesis has provided the following novel contributions to the field of predictive processing:

1. Cross-modal predictive information, due to extensive training, facilitates conscious access.
2. I Identified a new type of temporal regularity learning, reflected by the perceptual echo and placed the results within a predictive processing framework.

3. I identified a neural correlate of temporal prediction based on duration processing that was shared between visual and auditory sensory predictive processing.

Chapter 2
Cross-modal prediction changes the timing of conscious access during motion-induced blindness

2.1 Introduction

As mentioned in Chapter 1, the increasingly influential framework of 'predictive processing' (PP) posits that the brain continuously generates and updates predictions about incoming sensory signals(Clark, 2012; Karl Friston, 2005; Seth, 2014)(Clark, 2012; Karl Friston, 2005; Seth, 2014) (Clark, 2012; Karl Friston, 2005; Hohwy, 2013; Seth, 2014) according to principles of Bayesian inference. Accumulating empirical evidence suggests that perceptual predictions or expectations strongly influence conscious perception. For example the perceptual hysteresis effect shows that prior knowledge can enhance and stabilise conscious perception: previously perceived stimuli can bias conscious perception to represent subsequent stimuli in the same form (Hock, Scott, & Schöner, 1993; Kanai & Verstraten, 2005; Kleinschmidt, Büchel, Hutton, Friston, & Frackowiak, 2002; D. Williams, Phillips, & Sekuler, 1986).

A central aim of this thesis is to link predictive processing with conscious access. Recent evidence also suggests that prediction facilitates conscious access (Pinto, Gaal, Lange, Lamme, & Seth, 2015). Melloni and colleagues showed that thresholds of subjective visibility for previously seen degraded targets were lower than for novel degraded targets, with changes in threshold accompanied by a shift in a neurophysiological signature of conscious awareness to an earlier time point (Melloni, Schwiedrzik, Müller, Rodriguez, & Singer, 2011). Similarly, Lupyan and Ward (2013) found that visually presented objects preceded by congruent auditory cues (spoken words) were faster to break through continuous flash suppression than those preceded by incongruent cues. These results, and others, collectively suggest that prior knowledge can enhance and

accelerate conscious access. However, research on how he development of expectation influences this acceleration effect is still lacking.

The current study was designed to assess how the learning of prior expectations changes the timing of conscious access. I intensively trained participants on a cross-modal predictive relationship between an auditory tone and a colour change of a visual target, and they were tested on separate experimental sessions how the learned predictive association affected the timing of conscious access of the visual target. Training phases and test phases were interleaved in order to explore the temporal nature of the influence of cross-modal perceptual predictions on conscious access.

To precisely measure the timing of conscious access, I employed the well-known motion-induced blindness (MIB) paradigm (Y. Bonneh, Cooperman, & Sagi, 2001). In MIB, a peripheral visual target disappears and reappears from awareness periodically when it is presented superimposed on a rotating background pattern. Because disappearances and reappearances are clear-cut in MIB (as compared to, for instance, binocular rivalry), participants can reliably report the timing of conscious access to the target. Moreover, previous research on MIB has shown that subjectively unseen visual information can still be integrated and updated to form new object representations unconsciously (Mitroff & Scholl, 2005). Furthermore, recently, Wu and colleagues demonstrated that a transient visual change (a flashed ring) can influence the timing of conscious access of an unseen target (Wu, Busch, Fabre-Thorpe, & VanRullen, 2009). These findings speak to the persistence of representations of target stimuli during periods of subjective invisibility during MIB. These characteristics of MIB allows us to induce visual expectation, while the persisting unconscious visual processing is on going and to examine how learning of visual expectation influence the timing of conscious access of the unconscious visual information.

To establish visual expectation, I trained human participants to learn an audio-visual

cross-modal association. Previous studies on association learning have shown that a learned association can be reactivated when either a cue or a target is unconscious. For instance, after learning in classical conditioning, autonomic responses (e.g. changes in skin conductance) can be triggered without conscious knowledge of the CS-US contingencies (Hamm & Vaitl, 1996, 1996; Knight, Nguyen, & Bandettini, 2003; Morris, DeGelder, Weiskrantz, & Dolan, 2001; Morris, Ohman, & Dolan, 1999; Morris, Ohrnan, & Dolan, 1998; Whalen et al., 1998; M. A. Williams, Morris, McGlone, Abbott, & Mattingley, 2004). Moreover, unconscious visual processing has been found to be influenced by expectation from explicit cues. In a blindsight patient study, Kentridge et al. showed that an explicit valid spatial cue helped discriminate orientations of a target stimulus which were presented within the patient's blind visual field without those targets eliciting awareness (Kentridge, Heywood, & Weiskrantz, 2004). This result suggests that informative expectations from an explicit cue can modulate visual processing even when the visual information is processed unconsciously. Taken together, I predicted that the learned expectation elicited by a firm audio-visual association should influence the unconscious visual processing which is visually supressed by MIB.

Previous research has also shown that microsaccades can easily trigger target reappeance in MIB (Y. S. Bonneh et al., 2010; Hsieh & Tse, 2009; Martinez-Conde, Macknik, Troncoso, & Dyar, 2006). Therefore, microsaccades may cause measuring noise in the reappearance timing. To avoid potential microsaccades and minimise visual interference in our experiment, it is necessary to present the cue through a modality other than vision. In the current study, an auditory informative cue was presented to elicit cross-modal expectation. Studies have been shown that concurrent presented cross-modal information from tactile (Lunghi, Binda, & Morrone, 2010), olfactory (Zhou, Jiang, He, & Chen, 2010), and auditory stimuli (Y.-C. Chen, Yeh, & Spence, 2011; Conrad, Bartels, Kleiner, & Noppeney, 2010) can modulate the suppression time of binocular rivalry based

on the perceptual or semantic congruency between the cross-modal information and supressed visual contents. Therefore, the auditory cue used in our experiment can serve the function of inducing visual expectations without interfering with the stability of MIB state.

I hypothesised that conscious access to such representations could be influenced by cross-modal cues if strong associations were formed via extensive training. To test this hypothesis, I delivered an auditory cue while visual target was rendered subjectively invisible via MIB, and I then gradually changed the colour of the target to assess how the congruency between the cue and the visual sensory event influenced the timing of subjective target reappearance. Timing was compared with that from control trials in which the target was physically removed. In this way, I were able to quantify the effects of prediction on awareness and while minimizing the impact of other cognitive process such as attention and response bias. The main prediction was that this effect on conscious access can be modulated though the interleaved association training sessions across the whole experiment.

2.2 Methods

2.2.1 Participants

Participants were 26 healthy students from the University of Sussex (7 male, 18–31 years; mean age 23.15 years, normal or corrected-to-normal vision). All of them provided informed consent before the experiment and received £15 or course credits as compensation. Two participants were excluded from data analysis because they misunderstood the instructions as to key responses. The experiment was approved by the University of Sussex ethics committee.

2.2.2 Stimuli and Procedure

Stimuli were generated using the Psychophysics toolbox (Brainard, 1997) and all visual stimuli were presented on a Dell Trinitron CRT calibrated display (resolution 1048x768; refresh rate 100 Hz) with a black background. Participants sat 50 cm away from the monitor, using a chin rest. A linearized colour lookup table was used for gamma correction (γ = 2.2).

Participants took part in a total of 4 training blocks and 4 MIB (testing) blocks across two consecutive days. As shown in Figure 2-1, each participant completed four interleaved training and MIB block on each day as follows: (i) Training (200 trials); (ii) MIB (90 trials); (iii) Training (100 trials); (iv) MIB (90 trials).

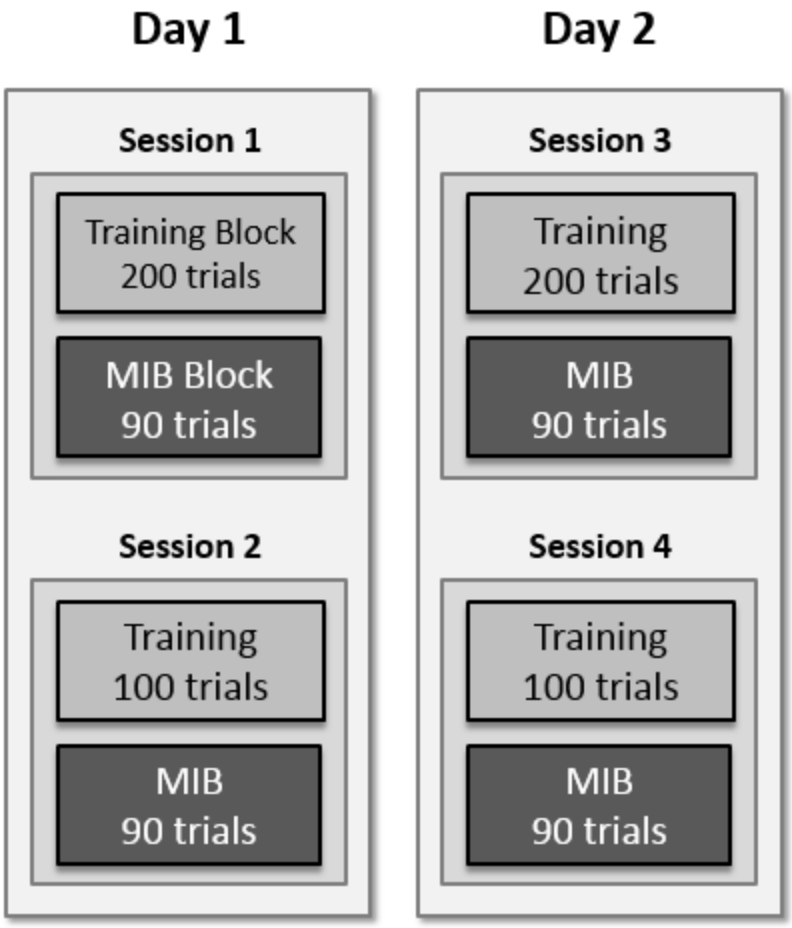

Figure 2-1: Organization of the experiment. Participants took part in two sessions on two consecutive days. Each session consisted of two 'training' and two 'MIB' blocks. Each MIB block consisted of 60 experimental trials interleaved with a block of 30 control trials.

2.2.2.1 The training procedure.

At the beginning of a training trial, participants were presented with a white fixation point and a blue target stimulus (Figure 2-2). The fixation point and the target were both circular dots subtending 0.2 degrees of visual angle. The fixation was presented at the centre of the screen and the target was to the upper-left of the fixation at 5.4 degrees of visual angle. Participants were

instructed to maintain their gaze at the central fixation point and to pay attention to the peripheral target during the entire trial. Following a delay (drawn from a random uniform distribution of 1-2 sec) an auditory cue (500 Hz or 1000 Hz pure tone) was presented for 300 ms. One second following the auditory cue onset, the target changed its colour (instantaneously) from its initial blue to either red or green. Crucially, the pitch of the auditory cue predicted the colour change with 80% validity. For example, after the 500Hz tone there was an 80% chance of a blue-to-red colour change and a 20% chance of a blue-to-green change. In this *training* task, participants were required to indicate the new colour of the target by pressing key '.>' (with the right index finger) or '/?' (with the right middle finger) as accurately and fast as possible. Feedback was presented on the screen after each correct and incorrect trial. The probability mapping between the cue and the colour change of the target, and the mapping between the colour changes and the response keys, were counterbalanced between participants.

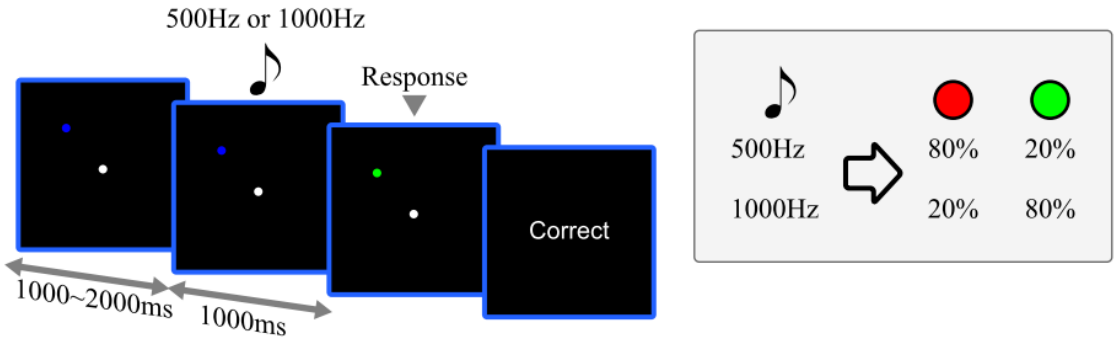

Figure 2-2: The trial sequence during training blocks. A fixation and a target (the blue dot) display were presented for 1,000 to 2,000 ms and followed by a 300 ms auditory cue (500Hz or 1000Hz). 1000 ms following the auditory cue, a colour change of the target occurred. Participants were instructed to respond to this colour change by pressing a key. The right panel illustrates the probabilistic relationships between the frequencies of the tone and the change of colours of the target.

2.2.2.2 The MIB procedure

In the MIB blocks (Figure 2-3a and Figure 2-3b), each trial also started with a central fixation point and a peripheral target. To induce MIB of the peripheral target, I presented a rotating pattern consisting of an array of 64 (8 x 8) grey crosses occupying 15.6ºx15.6º in the background. The speed of rotation was 0.5 cycle/s. The rotating background temporarily induced the subjective disappearance of the target, this being MIB. Participants were instructed to indicate the target disappearance and reappearance by pressing and releasing the 'z' key respectively (with the left index finger). Immediately after the report of disappearance, the colour of the target started changing gradually and linearly in RGB space from blue to green or red. The complete colour change took 3 seconds. While this colour change was occurring, 1 second after target disappearance, a 500 Hz or 1000 Hz tone was delivered to the participants. The participants were instructed to ignore the tone and to release the key as quickly as possible following subjective reappearance of the target. The target remained on the screen for 300 ms after the key was released. This was followed by a colour judgement task in which participants had to report the new colour of the target with the same response keys as in the training block. In cases where participants released the key before the colour change had completed, they were instructed to indicate whether the colour was closer to red or green. This response ended the trial. If participants released the key (signifying subjective target reappearance) before auditory cue presentation, the trial was discarded. The probabilistic relationship between the pitch of the tone and the change of the colour remained the same as the training procedure (80%). The time to reappearance, defined as the duration between the onset of the cue and reappearance of the target reported by key release, was used to measure the influence of prediction on the timing of conscious awareness of the target. Before the first MIB block started, participants performed

several practice trials (> 20) to learn how to acclimatise to the MIB procedure in order to ensure they reliably experienced long periods of target invisibility.

To exclude the effects of attentional bias and congruency effects at a response stage, participants undertook control trials in which the target was physically removed from the stimulus display (Figure 2-3c). Participants followed the same instruction in the control trials as in experimental trials: They needed to indicate the disappearance and reappearance of the target and make a subsequent target colour judgement. To match the individual MIB behavioural profiles (see Figure 2-4 for individual distribution of MIB disappearance duration), control trials were designed as 'replays' of experimental trials within the same block: Durations of subjective invisibility and reappearance colours were sampled from the within-block experimental trials. Each sampled experimental trial was used to construct four congruent control trials and one incongruent control trial. These control trials were randomly inserted into the remaining period of the same block. The duration between the onset of a trial and the onset of the physical disappearance was randomly selected between 2 to 4 seconds for each control trial. Each MIB block consisted of 60 randomly interleaved experimental trials and 30 control trials.

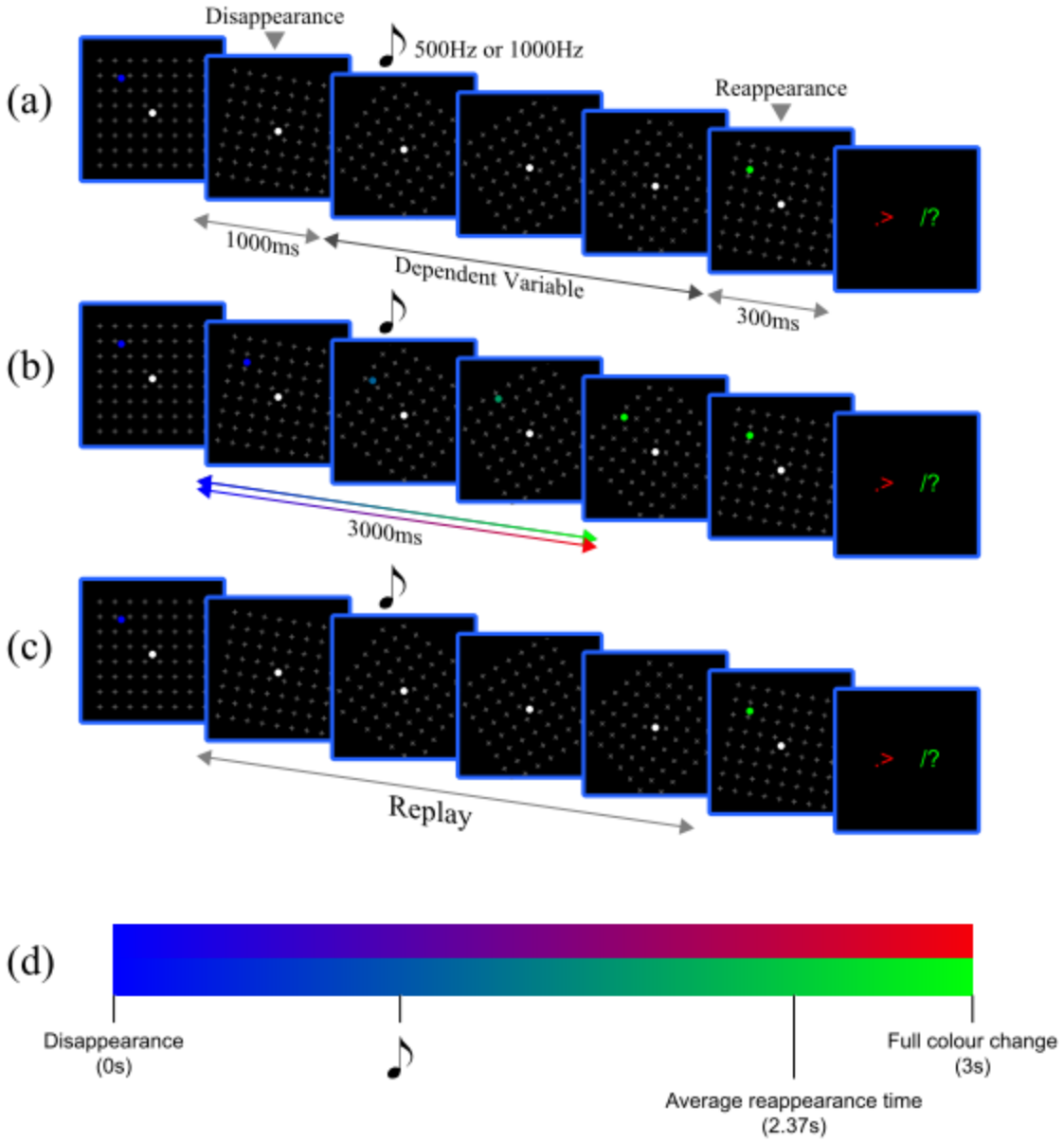

Figure 2-3: The trial sequence during motion-induced blindness (MIB) blocks: (a) the subjective perceived sequence (b) the experimental sequence. (c) the control sequence. Each MIB experimental trial (a, b) started with a central fixation point, a peripheral target, and a constantly rotating background. The rotating background temporarily induced the subjective disappearance of the target and participants were instructed to indicate subjective disappearance and reappearance times by pressing and releasing a key. After the key was pressed (i.e., the onset of subjective invisibility), the colour of the target started changing gradually from blue to green (or red); the complete change lasted 3 seconds. 1 second after the start of this colour change, a 500 Hz or 1000 Hz tone (the cue) was delivered. The reappearance duration, defined as the duration

between the onset of the cue and the key release was used to estimate the influence of the prediction on the conscious awareness of the target. At the end of a trial, participants were instructed to indicate whether the colour was more similar to pure red or pure green. In a control trial (c), participants followed the same instruction in as in experimental trials. However, the target was physically removed from the stimulus display. (d) The time course of colour change in the experimental sequence. Note that the colours are highly distinguishable at the (average) reappearance time.

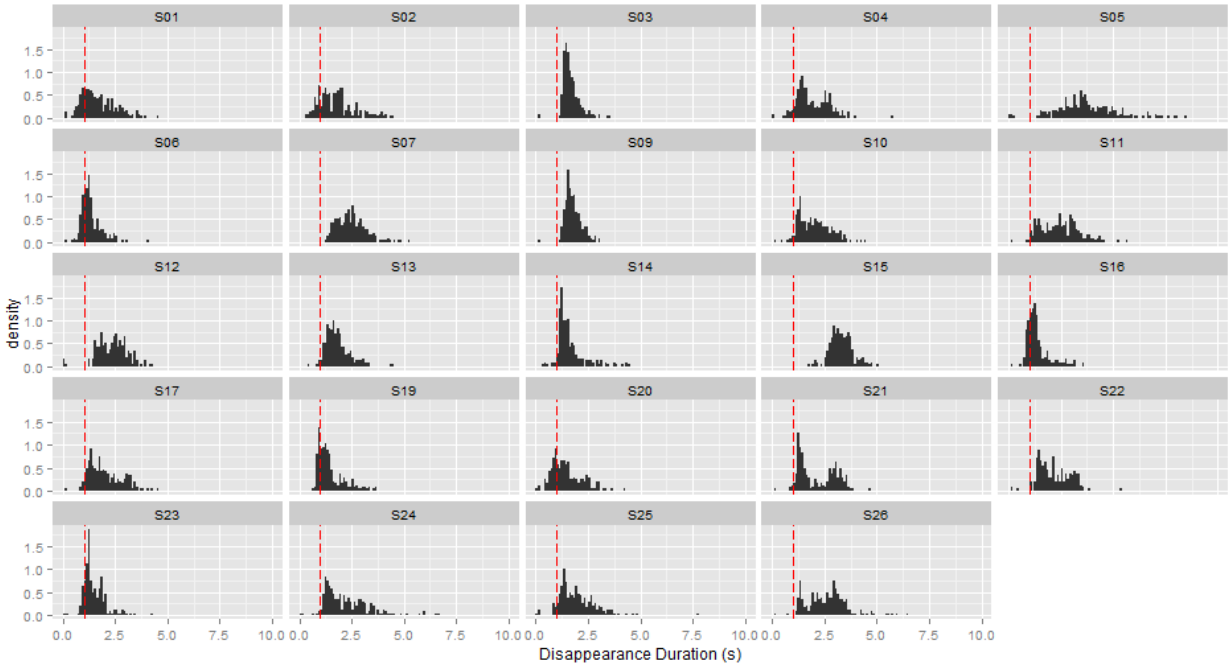

Figure 2-4: Distributions of time-to-reappearance from individual participants. There are substantial individual differences in the timing of perceptual disappearances in MIB. To match the individual MIB behavioural profiles for control trials, I 'replayed' experimental trials as control trials within the same block: Durations of subjective invisibility and reappearance colours were sampled from the within-block experimental trials. Therefore, the timing distribution of physical disappearance in the control trials resembles disappearance duration the distribution of experimental trials. This design ensures our main results should not be affected by these individual differences. Red lines indicates the timing of the sound delivery.

After completing the fourth MIB block, participants completed a brief questionnaire

concerning the subjective evaluation of the influence of the auditory cue on the disappearance duration of the visual target (a 5-point-scale question “Do you feel that the pitch of the sound influenced your perception of the target during the MIB state? From “1-not at all” to “5-I strongly felt the influence”).

2.3 Results

2.3.1 Training Blocks

In the training blocks, analysis of response times (RTs) indicated that participants successfully learned the tone-colour probabilistic association. I defined the blockwise *congruency effect* (dependent variable) as the mean RTs of incongruent trials minus the mean of congruent trials, for a given block. The congruency effects of each block were submitted to Bonferroni-corrected one-tailed one-sample t-tests. I found that all congruency effects were significantly higher than 0 (all corrected $ps < 0.05$, see Figure 4). This indicates that participants learned the statistical relationship between the cue and the target in all training blocks. The mean accuracy in the training block was 97.65% (SD = 2.03%) and all participants showed above 90% mean accuracy. The high accuracy indicates participants correctly performed the task.

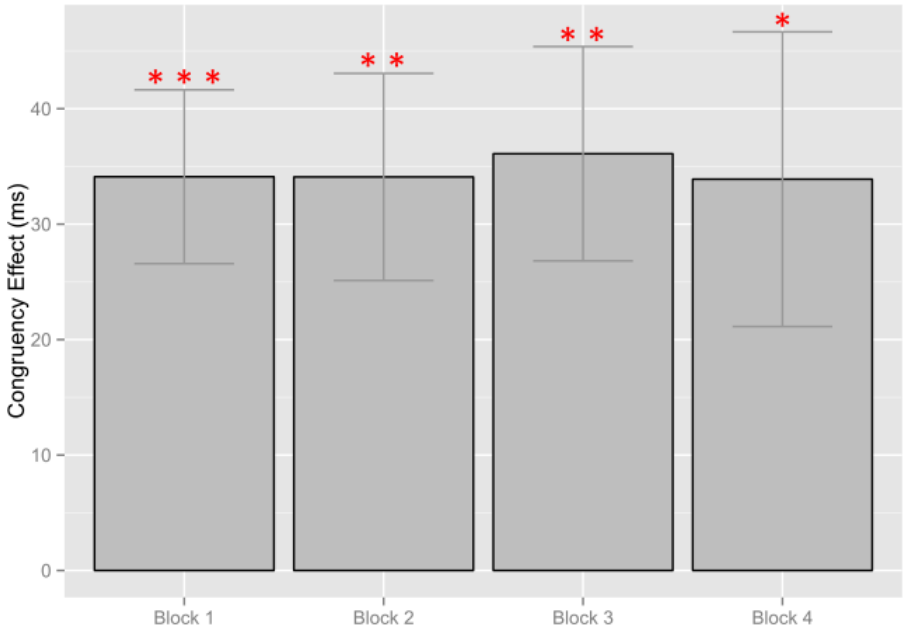

Figure 2-5: Response times (RTs) in the training blocks. The congruency effect was defined as the mean RTs of incongruent trials minus the mean RTs of congruent trials. Bonferroni corrected one-sample t-tests indicate that the congruency effect was significantly greater than zero in all the four training blocks. This indicates that participants learned the statistical relationship between the cue and the target at the very beginning of the experiment and maintained the learned association across the four training blocks. Error bars represent one standard error of the mean across participants. (P < .05, ** P < .01, *** P < .001)*

Figure 2-6 shows the reaction time data separated by every 50 trials in the training blocks. The congruency effect was significant (*ps* < .05, uncorrected) except the first 50 trials in the first block and the last 50 trials in the fourth block. This result implies that participants learned the association after 50 trials. Our data shows the dynamics learning of probabilistic knowledge.

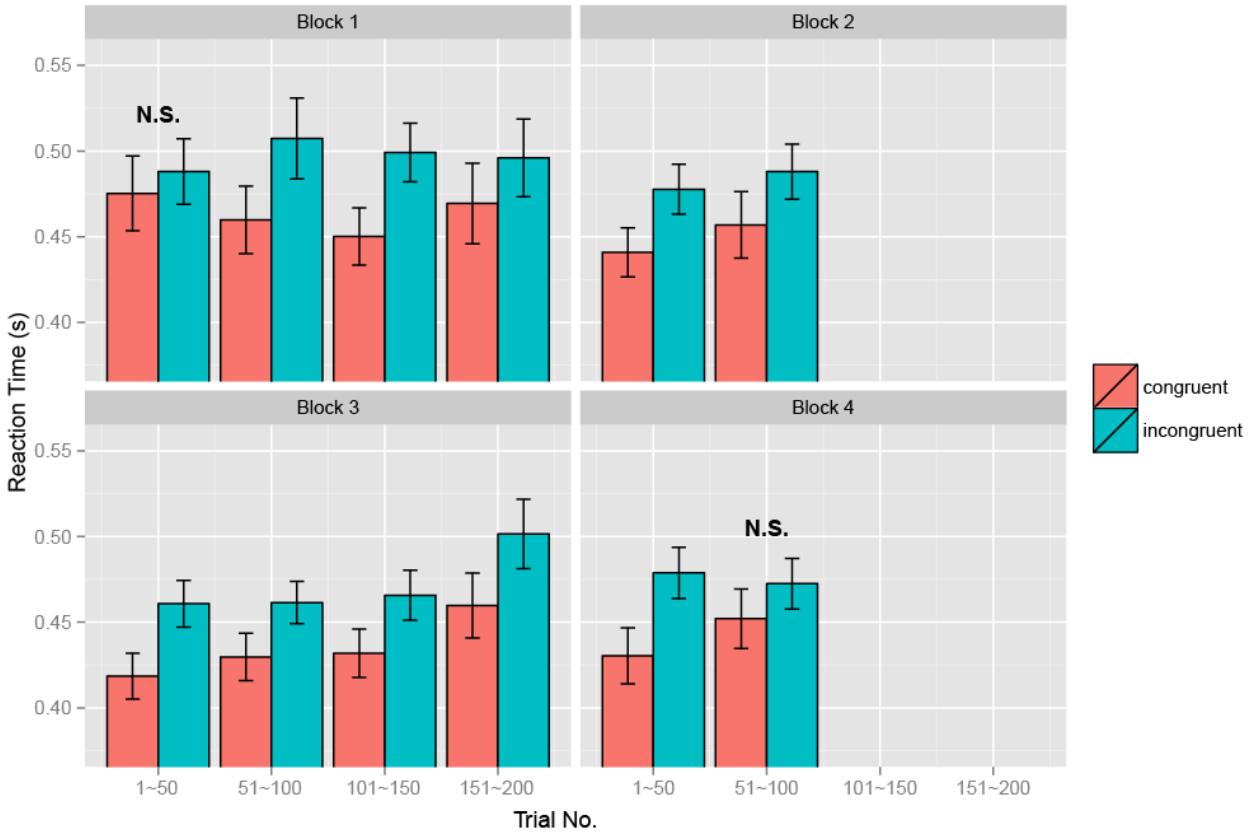

Figure 2-6: The reaction time data separated by every 50 trials in the training blocks. N.S.: No significant difference between conditions. Error bars represent one standard error of the mean across participants.

2.3.2 MIB Blocks

I next analysed the MIB blocks. The top panel of Figure 2-7 shows the mean time-to-reappearance in each condition. The overall time-to-reappearance in control trials was longer than in experimental trials (1780 and 1114 ms respectively, $t(23) = -14.78$, $p < .001$). This was expected due to the 'replay' nature of the control trials, which ensures that the response time for the experimental trials provides a lower bound on the response times possible in the control trials (excluding errors).

To test if the learned associations modulated timing of conscious access in the MIB blocks, I

similarly defined the blockwise congruency effect as the mean time-to-reappearance of incongruent trials minus the mean time-to-reappearance of congruent trials. First, I found significant differences across block (one-way repeated-measure ANOVA with the four-level factor Block, $F(3, 69) = 3.20$, $p = .03$, $\eta^2 = .12$). Next, polynomial contrasts revealed a strong linear trend in congruency effect from block 1 to block 4, $F(1, 23) = 6.91$, $p = .015$, $\eta^2 = .23$ (Figure 2-7 lower panel). Post hoc analyses revealed a significant difference between block 1 and 3, $t(23) = 2.90$, $p < .01$, and block 1 and 4, $t(23) = 2.34$, $p = .028$. These results indicate that the congruency effect in the MIB condition increased with successive blocks of training. No significant congruency effect was found by ANOVA and polynomial contrast analysis in any control blocks (all $ps > 0.5$). Thus, the learned cross-modal associations significantly accelerated conscious access, with this influence reliably increasing over time.

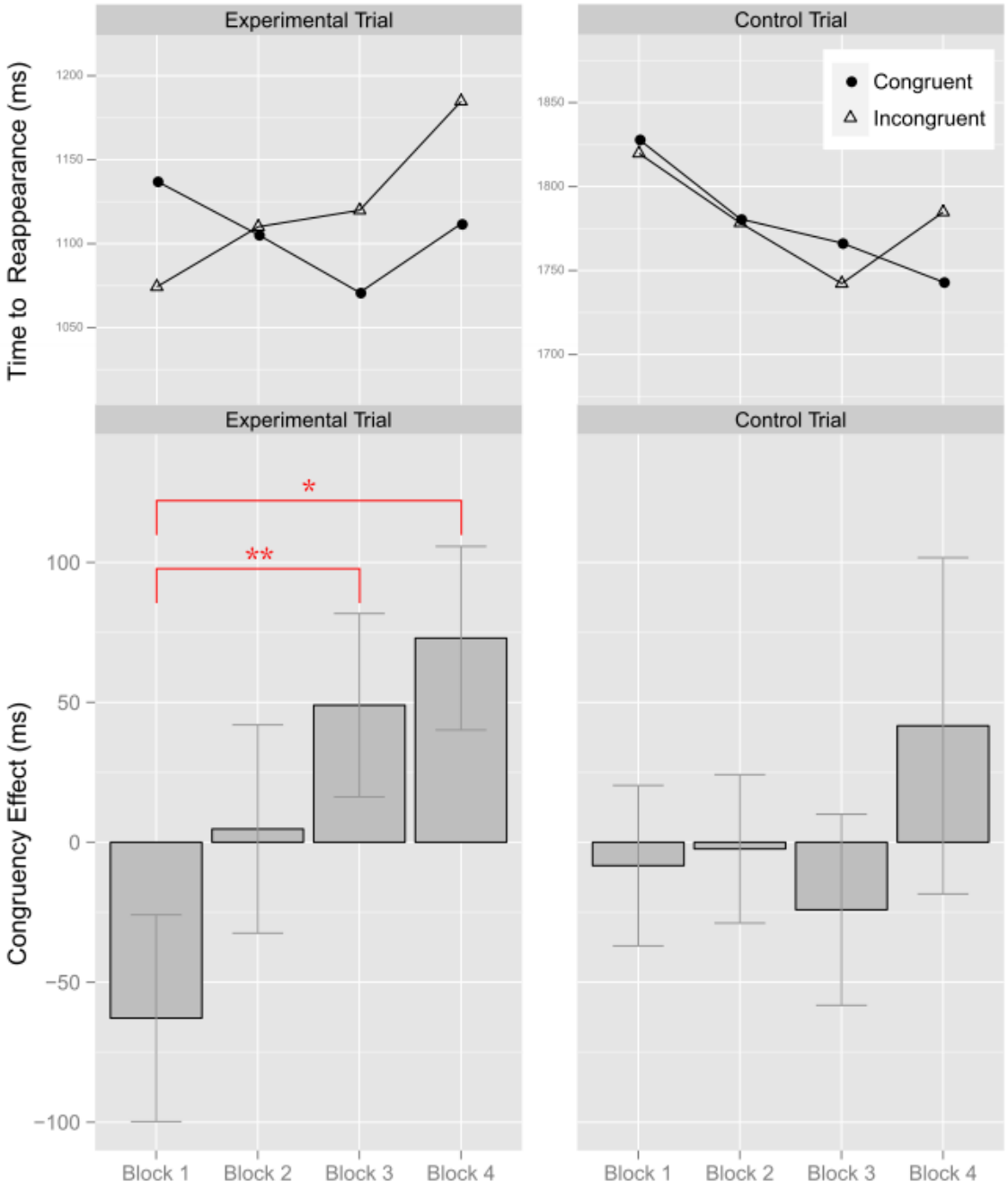

Figure 2-7: The upper panel show mean time-to-reappearance in each condition, and by block. The lower panel shows the corresponding congruency effects. The congruency effect showed a linear trend increment from Block 1 to Block 4 for experimental trials (left) but not for control trials (right). Post-hoc t-tests revealed that congruency effects in Block 3 and 4 were significantly larger than in Block 1. Error bars represent standard errors. (P < 0.05, ** P < 0.01)*

To assess that the results were not driven by outlying trial, I tested the congruency effect by

defining the blockwise congruency effect as the median time-to-reappearance of incongruent trials minus the median time-to-reappearance of congruent trials. First, one-way repeated-measure ANOVA with the four-level factor Block reached significance, $F(3, 69) = 3.20$, $p = .03$, $\eta^2 = .12$. Next, polynomial contrasts revealed a linear trend in congruency effect from block 1 to block 4, $F(1, 23) = 10.32$, $p = .04$, $\eta^2 = .31$. Finally, Post hoc analyses revealed a significant difference between block 1 and 3, $t(23) = 3.17$, $p = .004$, and block 1 and 4, $t(23) = 3.0$, $p = .006$. No significant congruency effect was found by ANOVA and polynomial contrast analysis in any control blocks (all *ps* > 0.6). Therefore, our results were not biased by outliers.

To ensure the results were not driven by outlying participants, I tested whether the congruency effect change was consistent across participants by adopting a general linear model (GLM) approach to estimate congruency effect at the individual level. For each participant, a GLM with one regressor (congruent versus incongruent) was applied to each block to estimate the congruency effect. The standardised regression coefficient of each block and each participant (i.e., beta) was computed and entered into the group level analysis as a dependent variable. The results of this analysis are highly comparable to the results computed by the mean time-to-reappearance difference mentioned above. First, I found significant differences across block (one-way repeated-measure ANOVA with the four-level factor Block, $F(3, 69) = 4.4$, $p < .01$, $\eta^2 = .161$, see Figure 2-8). Next, a polynomial contrast performed on betas revealed a strong linear trend in congruency effect from block 1 to block 4, $F(1, 23) = 11.26$, $p < .01$, $\eta^2 = .329$. Post hoc analyses revealed significant differences between block 1 and 2, $t(23) = 2.07$, $p = .05$, 1 and 3, $t(23) = 3.02$, $p < .01$, and block 1 and 4, $t(23) = 3.2$, $p < .01$. No significant congruency effect was found by ANOVA and polynomial contrast analysis in any control blocks (all *ps* > .3). The results indicate that the congruency effect change across blocks was reproducible at an individual-participant level and were not driven by outliers.

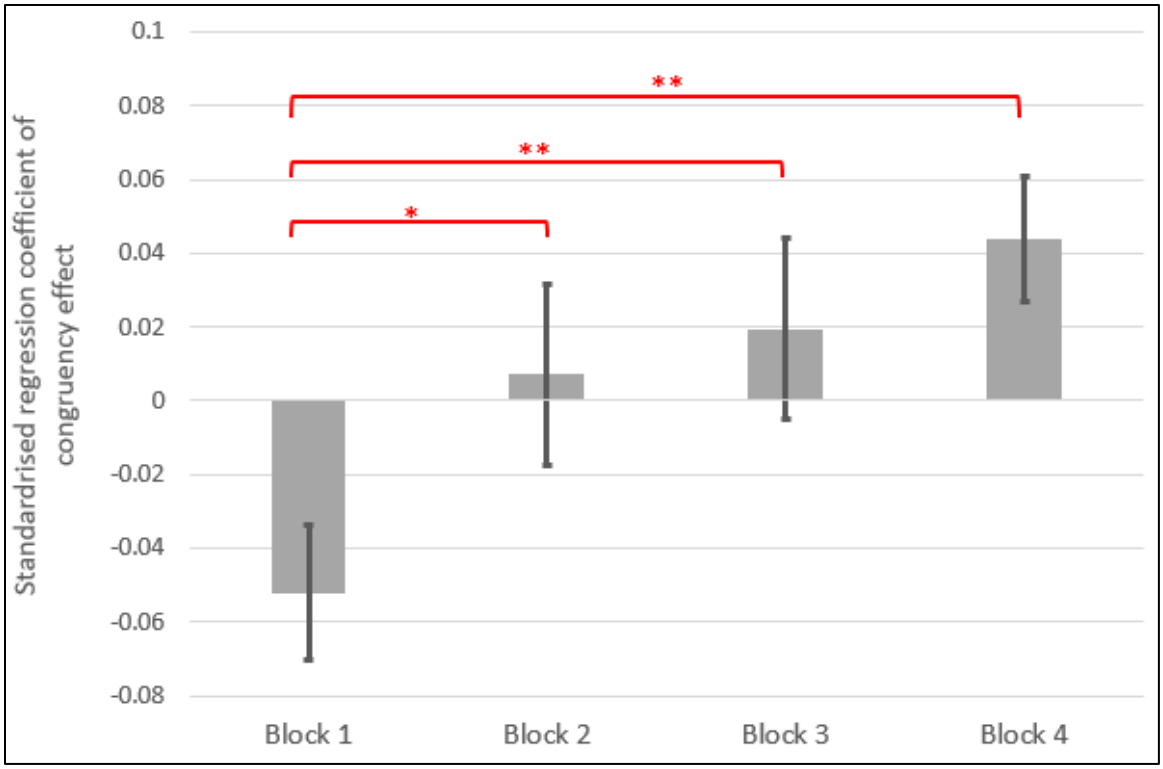

Figure 2-8: Congruency effect using the GLM approach. Error bars represent standard errors. (P < 0.05, ** P < 0.01)*

To further establish that these congruency effects in the MIB blocks were associated specifically with the tone-colour training, I performed an additional analysis. I reasoned that if the changes of the congruency effects from the first MIB block to the fourth block were influenced by the training, then the magnitude of changes in congruency effect over the training blocks should predict the magnitude of changes of congruency effect during the experimental sessions. To balance training trial numbers in this analysis, I calculated the magnitude of mean congruency effect in the first two blocks and last two blocks and the difference between the two means for both training and MIB trials, and then examined the correlation in the differences across participants. Results confirmed our prediction. Applying a Spearman's rank correlation test, I found a significant correlation (*rho* = .45, p = .026, see Figure 2-9) indicating a linear relationship between the change of association strength and the change of influence of prediction on conscious

awareness as indexed by congruency effect. There was no significant correlation when the same analysis was applied to control trials (*rho* = -0.16, *p* = .43). I also performed this Spearman's rank correlation analysis with outliers excluded data (by a 3 SD criterion and one data point was excluded). The result shows marginal significance, very close to the significance level (*p* = .052). This suggests that the observed correlation was not driven solely by outliers but veridically driven by the trend across all participants.

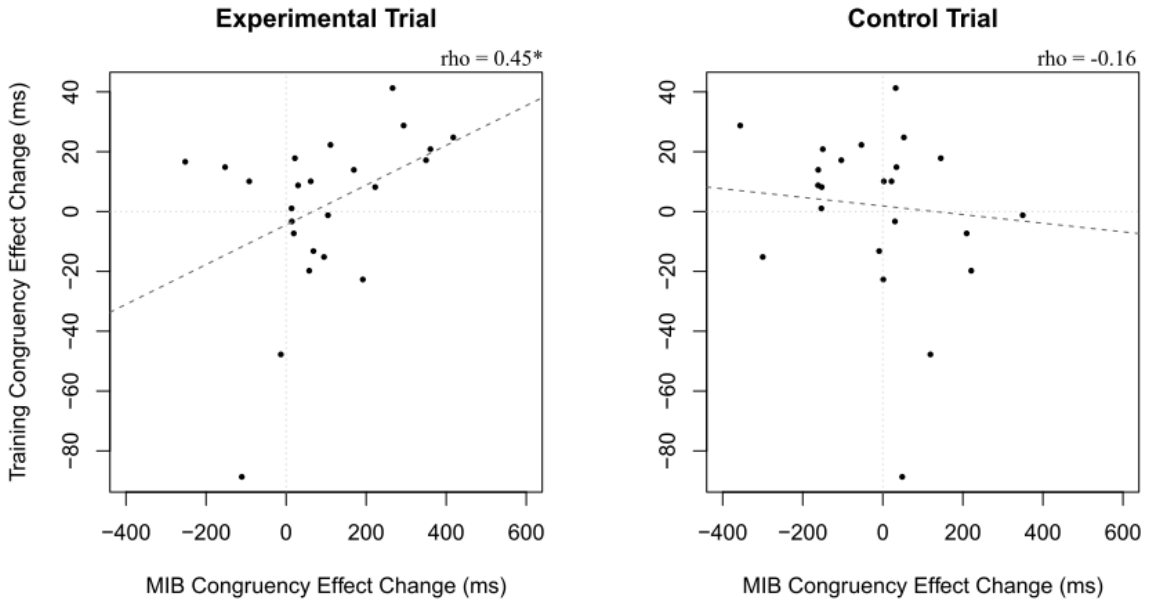

Figure 2-9: A Spearman's rank correlation test on MIB congruency change and the training congruency effect change between the first (block 1 and 2) and the second half (block 3 and 4) of the experiment. Data from the experimental trials showed a significant correlation. No effect was found in the control trials.

To exclude any possible confound due to problems in discriminating target colour on reappearance, I computed the mean accuracy of the colour judgement in the MIB. The accuracy was nearly perfect (99.23%, SD = .7%) indicating that participants did not have difficulty in

discriminating the colour upon reappearance. Due to the ceiling effect, I did not observe any significant differences in accuracy between the congruent and incongruent conditions in both experimental and control trials (all p-values > .38).

Finally, to test whether participants were aware of the influence of prediction on the time-to-reappearance, I computed the correlation between the self-evaluation scores from the post-experiment questionnaire and the congruency effects. I did not find a significant relationship between them (r = -.253, p = .233), suggesting that participants' self-evaluation of the influence of the pitch was independent from the actual strength of influence from the predictive cues.

2.3.3 Colour Vision Assessment

In the current study, I did not conduct a formal assessment for colour vision. However, the results of our training and MIB trials ensure that our participants had the ability to discriminate the colours used in our present study. The Figure 2-10 shows the average accuracy from all participants from the training blocks and MIB blocks. All participant performed above 90% accuracy in both blocks. This result provides evidence that all participants had the colour vision needed to perform these tasks.

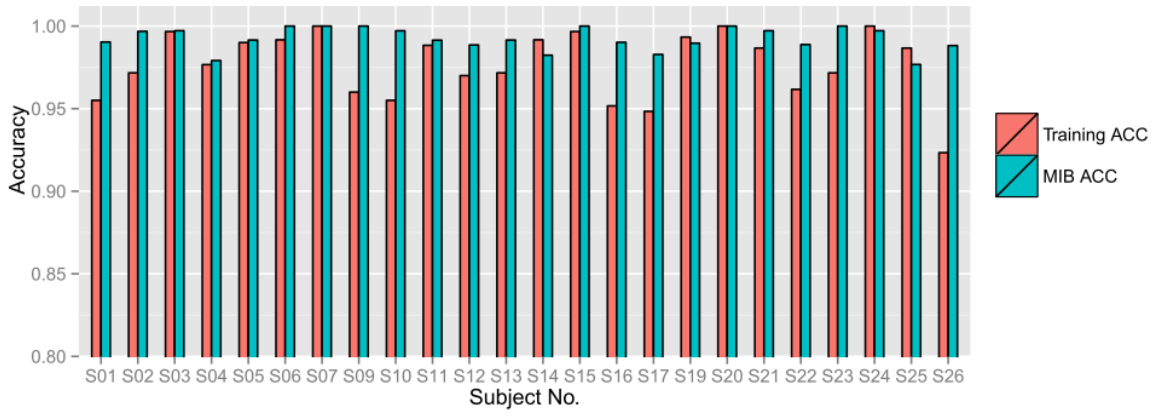

Figure 2-10: Average accuracy of colour judgments from individual participants in the training and MIB blocks. All participant performed above 90% accuracy in both blocks suggesting all participants had the colour vision needed to perform these tasks.

2.3.4 Distributions of time-to-reappearance by blocks

One potential confound of our main finding of congruency effect increasing by block is that participants may have performed the task better over time. Participants may have been able to maintain fixation more steadily and increase the time-to-reappearance longer. Therefore, in the later blocks, expectation may have had more chance to influence visual processing with more practice and enlarge the congruency effect. However, according to this account, the lack of congruency effect in earlier blocks was not due to weak learned association in the earlier stage of the experiment. Alternately, it may have been caused by larger variance and shorter duration of the MIB state. To rule out this potential confound, I compare the distributions of time-to-reappearance between blocks. Figure 2-11 shows the distribution of time-to-reappearance. Pairwise Kolmogorov-Smirnov tests showed no significant difference between distributions (Bonferroni-corrected $ps > .05$). Therefore, the difference of congruency effect between blocks was not driven by the time-to-reappearance profiles in different blocks.

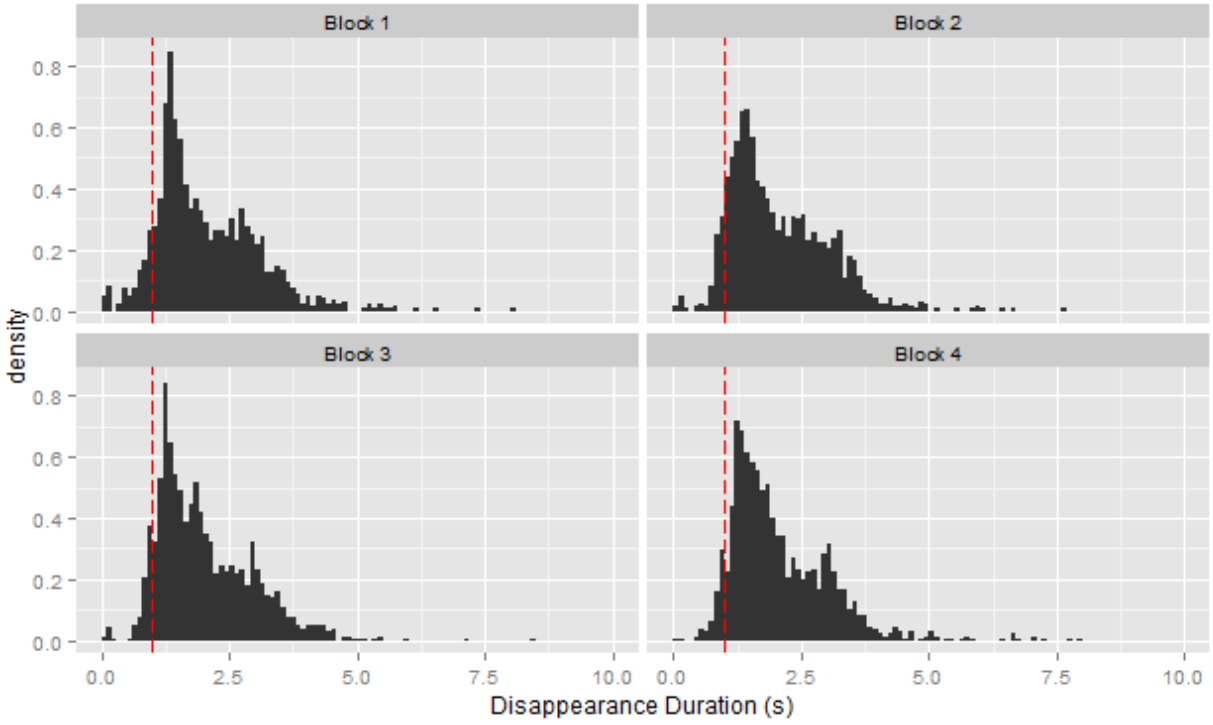

Figure 2-11: Distributions of time-to-reappearance in each MIB blocks. The red lines indicate the onset of auditory stimuli. Pairwise Kolmogorov-Smirnov tests showed no significant difference between distributions (Bonferroni-corrected ps > .05), suggesting that the difference of congruency effect between blocks was not driven by the time-to-reappearance profiles in different blocks.

2.4 Discussion

This chapter has described psychophysical data showing that learned cross-modal associations influence the timing of conscious access during motion-induced blindness. In an interleaved training-testing design I trained participants to learn associations between tones and colours, and then – using tones as predictive cues – explored whether congruent tone cues modulated the subjective timing of reappearance of coloured target stimuli. The results showed that cross-modal predictive cueing accelerated time to conscious access, but only after extensive training. I found a linear relationship between the amount of training and the degree of facilitation,

further underlining that learning of predictive associations can regulate their influence on the timing of conscious access. This finding was further supported by a correlation between the individual improvements in prediction learning from day one to day two, with the magnitude of the prediction effect across the same time period.

Crucially, the current design allowed us to exclude that the effects of prediction on the timing of conscious access were due to motor preparation or feature-based attention. First, the procedures and task goals were different for training and MIB blocks. In the training blocks, the task was to detect the colour change of target and to respond by pressing one of two keys to report the colour. In the MIB blocks, the task was to detect the reappearance of the target and to respond via a single key press. This difference in task makes it unlikely that participants adapted a similar motor preparation strategy acquired during the training to the MIB blocks. Second, it is well known that feature-based attention can facilitate the visual processing and detection performance of attended target features (Maunsell & Treue, 2006). One might wonder whether the prediction effect I found in the MIB blocks could be attributed to the facilitation of feature-based attention triggered by the auditory cue to a specific target colour. However, the control trials exclude this possibility, because the control trials were designed to replicate the subjective experience of the experimental trials but with physical disappearance and reappearance of the target from the stimulus array. In both experimental and control trials, participants performed the same task (detection of target reappearance). Crucially in these control trials, I found no difference in subjective time-to-reappearance between congruent and incongruent trials. Therefore, feature-based attention cannot account for the congruency effects of predictions on the timing of conscious access in this paradigm. In sum, these results taken together support that learned cross-modal predictive associations were responsible for the shift in the timing of conscious access.

One important innovation of this study is the demonstration of the flexibility that *arbitrary*,

cross-modal predictions can influence the timing of access to subjectively suppressed visual information through training. Previous studies focus on the prediction between long-term bound representations, e.g. semantic-object identity (Y.-C. Chen et al., 2011; Lupyan & Ward, 2013) or audiovisual speech (Alsius & Munhall, 2013; T. D. Palmer & Ramsey, 2012a) and motion integration (Conrad et al., 2010). For example, Conrad and colleagues (Conrad et al., 2010) show that, in binocular rivalry, motion sounds can stabilise and lengthen dominance periods for congruent visual motion. Similarly, Chen, Yeh, and Spence found that auditory cues shortened the suppression period of congruent visual objects. Taken together, these findings suggest that cross-modal predictive information exerts substantial influences on unconscious processing, conscious perception, and their interaction.

It is known that saccadic eye movements during MIB can cause immediate target reappearance. Previous research have also shown that microsaccades can modulate target perceptual transitions (Y. S. Bonneh et al., 2010; Hsieh & Tse, 2009; Martinez-Conde et al., 2006). Based on these previous findings, one might wonder whether these results might be explained by eye movements occurring preferentially when colour changes were congruent with auditory cues. I first note that participants were asked to maintain fixation during each trial, so saccades could be expected to be rare. In addition, in this experiment, colour changes happen both gradually and during perceptual suppression by MIB, making it is highly unlikely such eye movements could be triggered in such specific manner. Additionally, during data analysis, I discarded those trials in which the target reappearance *preceded* the predictive sound onset, where plausibly saccadic effects could have caused target reappearance. To sum up, the current result was unlikely caused by any impact of eye-movement.

What implications do the current results have for possible neural mechanisms? One possibility is that cross-modal predictions modulate the strength of low-level cortical stimulus representations.

Recent findings by Kok, de Lange, and colleagues show that learned auditory to visual predictions can create a neuronal template of the expected visual stimulus in the visual cortex (Kok, Failing, & de Lange, 2014) and bias visual representations as a result of the facilitation of information processing for the expected stimulus (Kok et al., 2013). This direct impact from cross-modal prediction to low-level visual representations may lead to a change in timing of conscious access, especially if these visual representations can be construed as a stimulus specific priors in a Bayesian framework. More generally these results strongly indicate the existence of neural pathways capable of conveying cross-modal predictive information. Previous studies have shown that, during associative learning, brain connectivity between cortical regions responsible for associated events increases in strength with training (Büchel, Coull, & Friston, 1999; den Ouden et al., 2009; Mcintosh & Gonzalez-Lima, 1998; von Kriegstein & Giraud, 2006). This suggests the possibility that, in the current study, the incremental change of the facilitation effect may be due to a gradual build-up of functional connectivity for conveying predictive information across the sensory modalities. Future investigations combining psychophysics with neuroimaging connectivity analyses could test this hypothesis.

In the current study, the congruency effect in the second day MIB blocks (the block 3 and 4) was much stronger than that in the first day MIB blocks (the block 1 and 2). A potential explanation for the second day 'boost' is that sleep between the two days facilitated and consolidated the associative learning from the first day. It is well known that sleep plays a beneficial role in memory consolidation (Stickgold & Walker, 2007; Walker & Stickgold, 2006) as previous studies have shown sleep facilitates associative memory (Stickgold, Scott, Rittenhouse, & Hobson, 1999), statistical learning (Durrant, Taylor, Cairney, & Lewis, 2011), and perceptual learning (Fenn, Nusbaum, & Margoliash, 2003; Karni, Tanne, Rubenstein, Askenasy, & Sagi, 1994; Stickgold, James, & Hobson, 2000; Yotsumoto et al., 2009). In the current study, participants learned the

statistical relationship between the auditory cue and the visual target. Given the extent literature on the role of sleep in learning, it is plausible that sleep strengthened the predictive value of the auditory cues, thereby facilitating conscious access to predicted visual events on the second day. However, in the current study was not explicitly designed to examine the effect of sleep in this experimental paradigm. Thus, I leave this possibility open for future studies.

The design of the current control trials is unable to rule out possible effect from response bias. The auditory cues may lower the response threshold of congruent visual targets. Therefore, in the experimental trial, when the visual targets are supressed in the MIB state, the evidence accumulation reaches the response threshold in congruent trials than in incongruent trials. However, in the current control trials, the visual target physically reappeared immediately avoiding the unconscious evidence accumulation process. Therefore, the control trials are unable to verify if the current congruency effect is led by the change of the response threshold. Future studies can make visual target reappear gradually allowing similar evidence accumulation processes to occur in control trial and examine the potential effect of change of response threshold.

The current results lend support to predictive coding (or predictive processing) frameworks for understanding perception (Clark, 2012; Karl Friston, 2005; Rao & Ballard, 1999). On these views, perceptual content is determined by probabilistic inference of the most likely external causes of sensory signals. The data support the view that conscious access occurs when predictive models are verified against sensory inputs so that prediction errors are minimized. On this view, valid predictive cues will engage predictive models that are validated against sensory signals more rapidly, leading to more rapid conscious access (Lupyan & Ward, 2013; Melloni et al., 2011; Pinto et al., 2015; Sherman, Seth, Barrett, & Kanai, 2015). The present results extend this approach by showing how predictive influences on perception develop across time via training, and by underlining the flexibility of these influences by demonstrating their efficacy using cross-modal

and arbitrary associations. These results together suggest that predictive influences may permeate and shape conscious experiences more deeply and broadly than previously thought.

Chapter 3
Visual Perceptual Echo Reflects Learning of Temporal Regularities in Rapid Luminance Sequences

3.1 Introduction

This Chapter aims to investigate how low level information within the sensory hierarchy is dealt with by predictive processing via unconscious visual sequence learning. However, relatively little is known about how the human visual system tracks, or learns about, such rapidly changing stimulus sequences. One recent finding relevant to this question is the 'perceptual echo' (VanRullen and MacDonald, 2012). This is long-lasting reverberation between a rapidly changing visual input and evoked neural activity, apparent in cross-correlations between occipital EEG and the stimulus itself. Specifically, when visually presenting a non-periodic dynamic sequence, whose luminance randomly fluctuated at a rate of 160 Hz, VanRullen and MacDonald found that the occipital EEG response displayed a periodic reverberation or 'echo' of the input sequence, which persisted for at least one second, and was found specifically in the alpha (~10 Hz) frequency range of the cross-correlation function, primarily over occipital electrodes and was observable at the group level.

Specifically, when presenting a random dynamic sequence, whose luminance randomly fluctuated at a rate of 160 Hz, the occipital EEG response displayed a periodic reverberation or 'echo' of the input sequence, which persisted for at least one second, and was found specifically in the alpha (~10 Hz) frequency range of the cross-correlation function, primarily over occipital electrodes. Importantly, the absence of such a reverberation when the luminance sequence of each trial was cross-correlated with EEG recorded on a different trial underlines that the perceptual echo is a true oscillatory response to the visual stimulation sequence, and not a general property of

ongoing EEG signals in response to this kind of stimulation. Furthermore, the seconds-long duration of the perceptual 'echo'– suggests a long-lasting representation of fast-changing sensory information over time. However, the functional relevance and the underlying mechanism of the perceptual echo have so far remained unclear.

One interpretation of perceptual echo is that the visual cortex repeatedly "replays" neural activation associated with afferent visual information. Supporting this view, studies in awake monkeys (Eagleman & Dragoi, 2012), mice and rats (Gavornik & Bear, 2014; Xu, Jiang, Poo, & Dan, 2012) found that after learning an association between a cue and a sequence stimulus, presenting the cue alone can elicit the same neural activation pattern as evoked by the actual stimulus. This replay activation pattern has been interpreted as a mechanism that may facilitate learning and memory consolidation (Euston et al., 2007; Skaggs and McNaughton, 1996).

I reasoned that, if the perceptual echo is indeed associated with visual regularity learning, the perceptual echo response should change across repeated presentations of the same luminance sequence, as the visual cortex encodes regularities across sequence repetitions. In two experiments, I tested this prediction by using random dynamic luminance sequences that were predictable across repeated presentations. In Experiment 1, I presented random dynamic luminance sequences that were each repeated four times and I calculated the perceptual echo for each presentation. In Experiment 2, I added an 'inverse' luminance sequence (which has the inverse luminance polarity of the original sequence) following the presentation of the fourth repetition of a sequence, which was followed by an additional presentation of the original sequence, to test whether changes in perceptual echo were specific to the (non-periodic) temporal and luminance information within a given sequence.

Altogether, this data provides first evidence for a predictive mechanism by which the human visual system rapidly learns temporal regularities in fast non-periodic sequences, and in doing so

provides a functional account of the alpha-band ‘perceptual echo’ response.

3.2 Experiment 1

3.2.1 Methods

21 healthy students participated in this study from the University of Sussex (11 male, 18–36 years; mean age 24.9 years, all had normal or corrected-to-normal vision). All of them provided informed consent before the experiment and received £10 or course credits as compensation for their time. The experiment was approved by the University of Sussex ethics committee.

Participants were seated in a dimly lit electromagnetically shielded room and their heads were stabilised in a head-and-chin rest to maintain centrality 50 cm away from a LaCie Electron blue IV 22" CRT Monitor at a 160Hz refresh rate, which had been manually gamma corrected. Stimuli were generated and presented using the Psychophysics toolbox (Brainard, 1997).

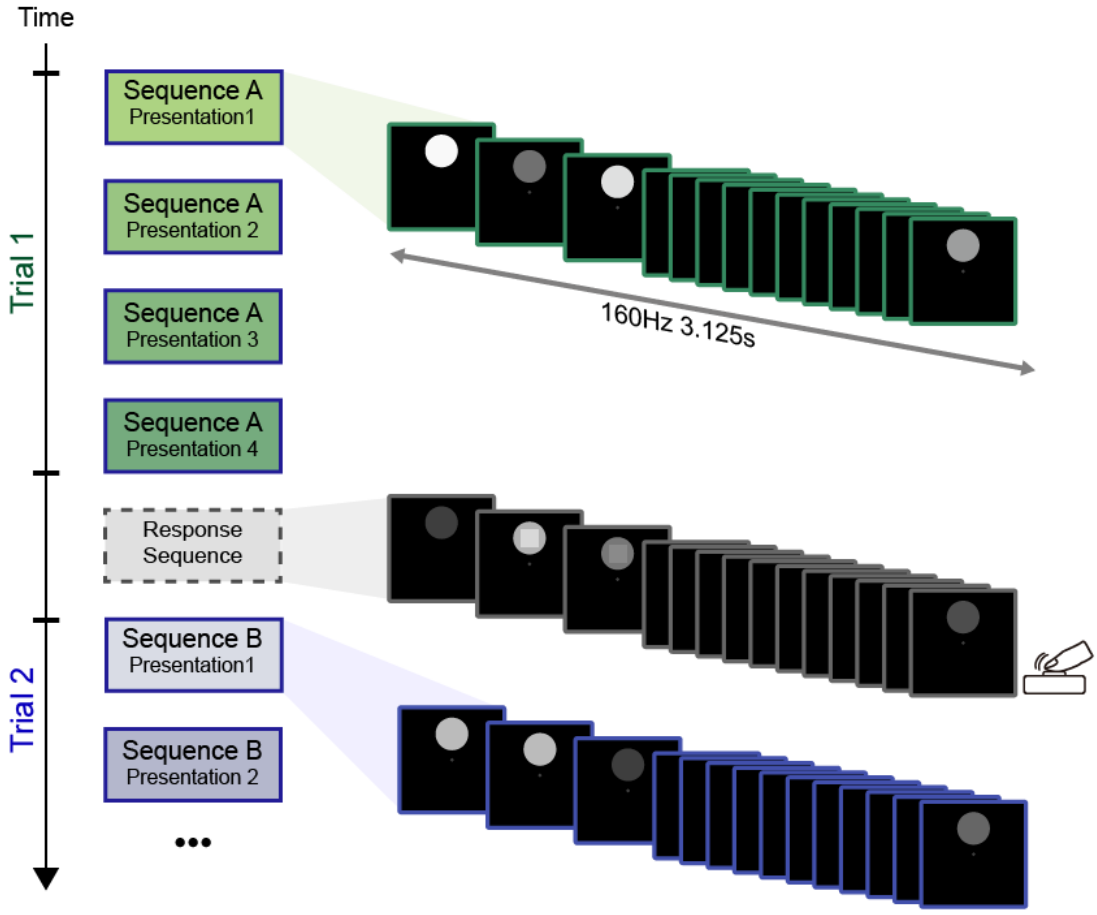

Figure 3-1: Example of trial structure in Experiment 1. For each trial a specific random luminance sequence (e.g. sequence A) was presented four times. Each luminance sequence lasted 3.125s and consisted of a disc with randomly changing luminance at 160 Hz. Sequences were separated by an inter-sequence-interval (ISI) of 3s. 'Response' sequences, which contained an embedded square image (1s duration, random onset time within sequence), were distributed at random points in-between trials throughout the experiment. Participants were requested to press a key after each sequence if they noticed a square.

Each experimental session comprised 60 trials utilising 60 unique random non-periodic luminance sequences. Each trial consisted of four presentations of the same sequence. Each presentation lasted for 3.125s and successive presentations were separated by an inter-sequence-interval (ISI) of 3s. Luminance-sequence stimuli were constructed based on VanRullen and Macdonald (2012) as

follows. Each sequence consisted of a disc stimulus subtending a visual angle of 7 degrees and its centre was placed at 7.5 degrees above a fixation point (a dim grey circle with 0.2 degrees of visual angle). The luminance of the disc altered randomly at a rate of 160 Hz, so that each sequence consisted of 500 luminance frames (see Figure 3-1). Thus, each sequence comprised a rapid non-periodic sequence of luminance changes. To ensure equal power of all temporal frequencies within each sequence, all sequences were processed by a whitening procedure. Fourier components of each sequence were obtained by fast Fourier transform. Power at all frequencies of each random sequence were equalised by normalizing the amplitudes of its Fourier components. An inverse Fourier transform was then applied to reconstruct the sequence. Thus, sequences were not distinguishable by power characteristics of their temporal frequencies.

To ensure participants maintained attention, sixty 'response sequences' were distributed throughout the experiment. Each 'response sequence' contained an embedded square image (3.75 degrees). The target appeared for 1s with onset time selected from a uniformly distributed random time during the sequence presentation. Participants were informed that the experiment was a visual detection task in which they were required to press the spacebar on a standard keyboard at the end of a sequence whenever they detected the target. Each response sequence was randomly assigned to a position in-between experimental trials. This was done individually for every response sequence, resulting in the possibility of there being one, more than one, or zero, response sequence between any two experimental trials. This design made it unlikely that participants could predict the onset of an experimental trial based on the occurrence of a response sequence.

Participants were not informed that there would be any repetition of the luminance sequences. The entire experimental session consisted of 300 (240 standard and 60 response) sequences and took approximately 1 hour to complete. In a post-experiment interview participants were asked if they noticed any repetition of luminance sequences in the experiment. The following questions

were asked:

Did you notice any experimental manipulation during this experiment?

Did you notice any relationship between any two sequences?

Did you notice any repetition of any of the sequences?

EEG data were recorded using a 64 channel ANT Neuro amplifier at a sampling rate of 2048 Hz. A 64 channel Waveguard EEG cap (ANT Neuro, Enschede) employing standard Ag/AgCl electrodes placed according to the 10-20 system was used. Horizontal and vertical eye movements were recorded using two independent electrode pairs. Impedances of recording electrodes were maintained below 10kΩ. No analogue filter was applied during on-line recording.

3.2.1.1 Data Processing and Statistical Analysis

Pre-processing and data analyses were performed using the EEGLAB toolbox (Delorme and Makeig, 2004) under Matlab (Mathworks, Inc. Natick, MA, USA), and custom Matlab scripts. The acquired EEG data were downsampled to 160Hz and filtered using a 2-80 Hz bandpass filter. Independent component analysis (ICA) was used to identify and remove ocular artefacts. Data in each trial were then epoched from 0 to 3.125 seconds time-locked to the stimulus onset (start of the luminance sequence).

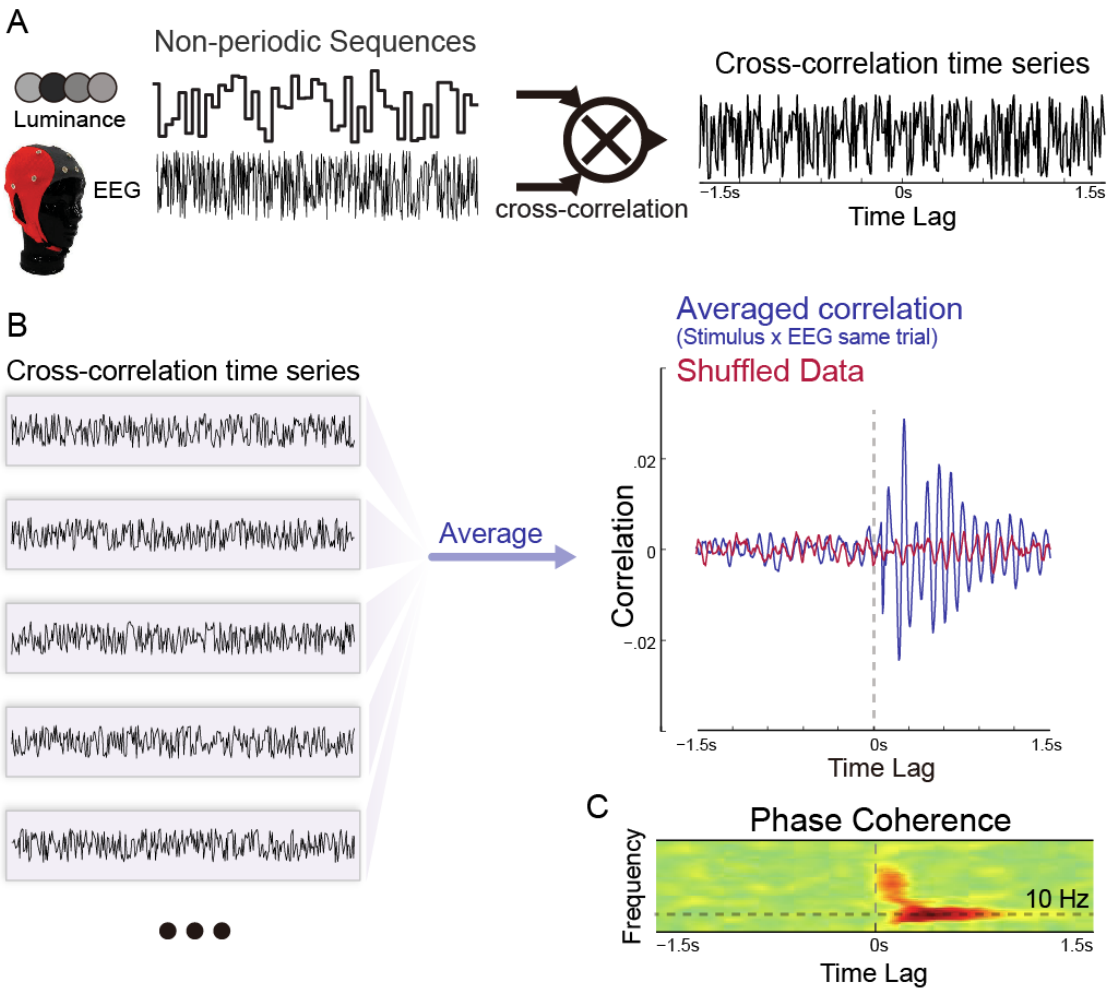

Figure 3-2: Recording and computation of perceptual echo for a representative subject. A. For each luminance sequence presentation and EEG sensor, I computed the cross-correlation (-1.5s:1.5s) between the luminance value series and the EEG time series. This provides an ensemble of cross-correlation time series for each luminance sequence, indexed by sensor and presentation number. B. As in VanRullen and Macdonald (2012), averaging the cross-correlations for each luminance sequence revealed a long lasting post-onset oscillation (~1.5 sec) in the alpha range: the 'perceptual echo' (blue line). The red line shows the same analysis with shuffled data (see methods), in which no perceptual echo is observed. C. The phase coherence (across trials) of the cross-correlation time series shows that the perceptual echo is due to the strong phase coherence within the alpha frequency range (~10 Hz).

To identify the perceptual echo, I calculated cross-correlations between EEG time series and

luminance time series for all experimental sequences and sensors. Cross-correlations were evaluated with lags between -1.5s and 1.5s (Figure 3-2a) by the following definition:

$$xcorr(t) = \sum_{T} stim(T).eeg(T + t)$$

xcorr is the cross-correlation value. *stim* and *eeg* denote the standardized stimulus sequence and the corresponding standardized EEG response, respectively.

For each participant, I then averaged the cross-correlations across each sequence for all sensors, which revealed a long-lasting post-onset oscillation in the alpha range (Figure 3-2b). This revealed a perceptual echo, replicating the results of VanRullen and Macdonald (2012). To quantify the amplitude of the echo response, I applied a Fast Fourier Transform on the average cross-correlation between 0s and 1s for each participant and sensor to extract the alpha-range (8-12 Hz) power.

To compare the echo response across successive presentations of a luminance sequence, the cross-correlations were averaged by the order of presentation for each sensor, across trials (e.g., all cross-correlations for the first presentation of a given random luminance sequence were averaged across trials, and the same for all second, third, and fourth presentations). I then computed the amplitude of the echo response for each presentation and participant.

To verify that the perceptual echo was driven by the EEG response to a specific luminance sequence and not by variations in ongoing alpha activity in the raw EEG signal, I created a 'shuffled' set of cross-correlations by randomly rearranging the EEG time series with stimulus sequences from different trials. As can be seen from Figure 3-2b (red line), this procedure leads to a complete absence of echo response, confirming that the echo response is not driven by the ongoing alpha-band EEG response.

I found that the magnitude of the echo response varied across individuals. Participants with no

significant perceptual echo were excluded from the data analysis, as follows. First, cross-correlation time series were averaged across all trials, then the distribution of absolute values across lags were derived from the averaged cross-correlation for real and shuffled data separately. The non-parametric Kolmogorov–Smirnov test was then performed to examine the similarity of the two distributions. I excluded those participants with $p > .001$ (i.e., those participants for whom echo responses were not significantly different between the shuffled and non-shuffled data). Seven out of twenty-one participants were excluded using this procedure (Figure 3-3).

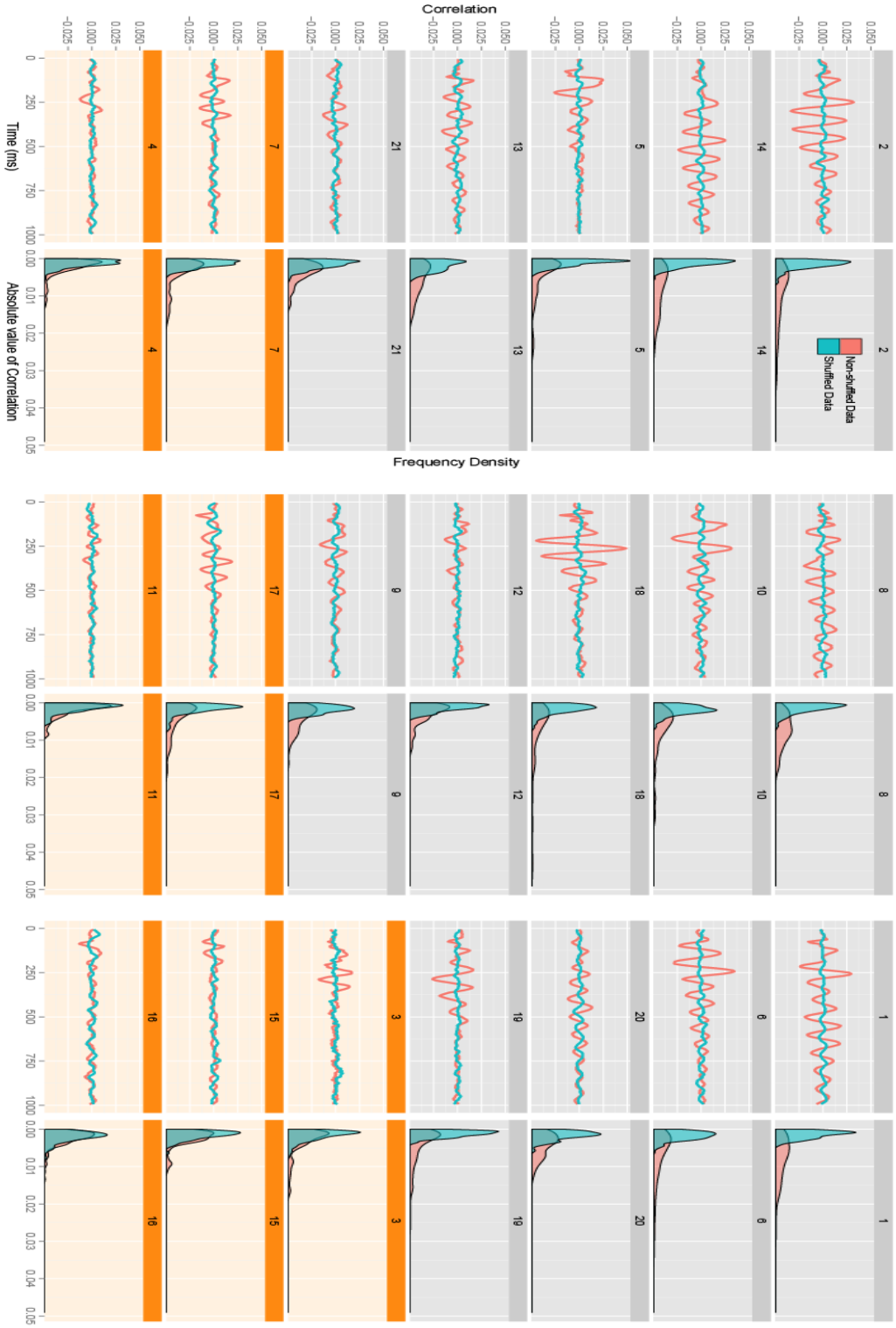

Figure 3-3: Individual echo response and its data distribution for Experiment 1. The magnitude of the echo response varied from individual to individual (the left panels in each column). To determine a rigorous exclusion criterion, cross-correlation time series were averaged across all trials (left panel). Distribution of absolute values across lags were derived from the averaged

cross-correlation (the right panels in each column) for non-shuffled (red) and shuffled data (blue). An average cross-correlation close to zero leads to a distribution skewed towards zero. The non-parametric Kolmogorov–Smirnov test was then performed to examine the similarity of the two distributions. We excluded those participants who showed the p-value of the comparison was larger than .001 (highlighted in orange). The order of participants is sorted by the result of the Kolmogorov–Smirnov test.

3.2.2 Results

Figure 3-4 shows the average alpha (8-12 Hz) power of the cross-correlation, i.e., the amplitude of the perceptual echo response, across all participants and sequence presentations. I found a maximal echo response over occipital sensors, centred over POz, consistent with previous findings (VanRullen & Macdonald, 2012). To maximize the sensitivity to any potential effect of stimulus sequence repetition on echo response, the following analyses were therefore conducted with data from POz only.

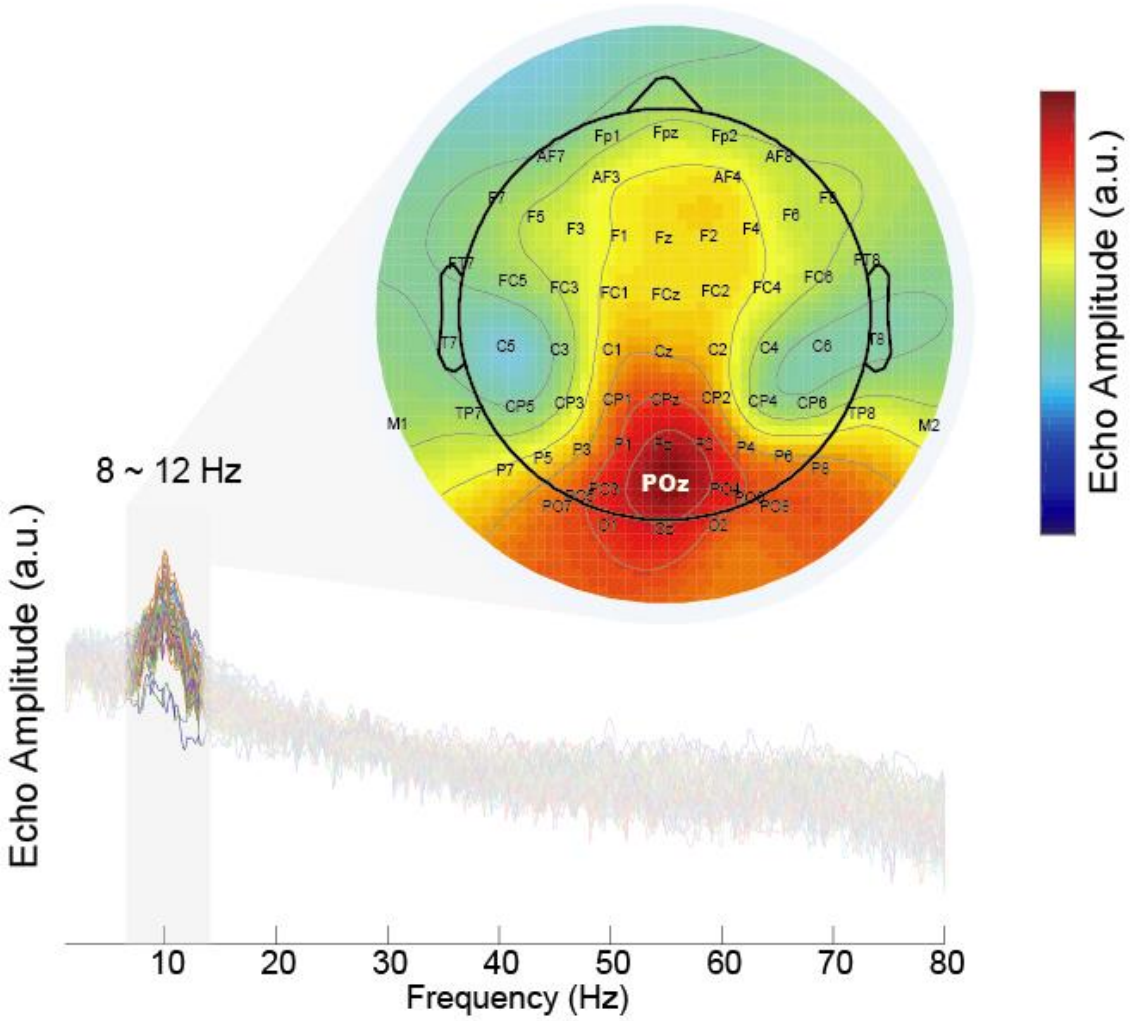

Figure 3-4: The topography of the echo response averaged across all sequence presentations and participants. Each line represents data from a single electrode. The topographic plot displays the 8-12 Hz envelope of the echo response and reveals a maximum at POz with a gradual decrease from posterior to anterior electrodes (Arbitrary Units (a.u.)).

To test whether repetitions of a luminance sequence modulated perceptual echo, I compared the average amplitude of the perceptual echo for each sequence presentation using polynomial contrasts. Supporting this hypothesis that successive presentations would increase echo amplitude, I found a strong linear trend from presentation 1 to presentation 4 ($F(1, 13) = 7.32$, $p = .018$).

Post-hoc paired t-tests revealed a larger echo amplitude for presentation 4 compared to presentation 1 (bootstrapping test with 10,000 resamples, $p < .01$), see Figure 3-5a.

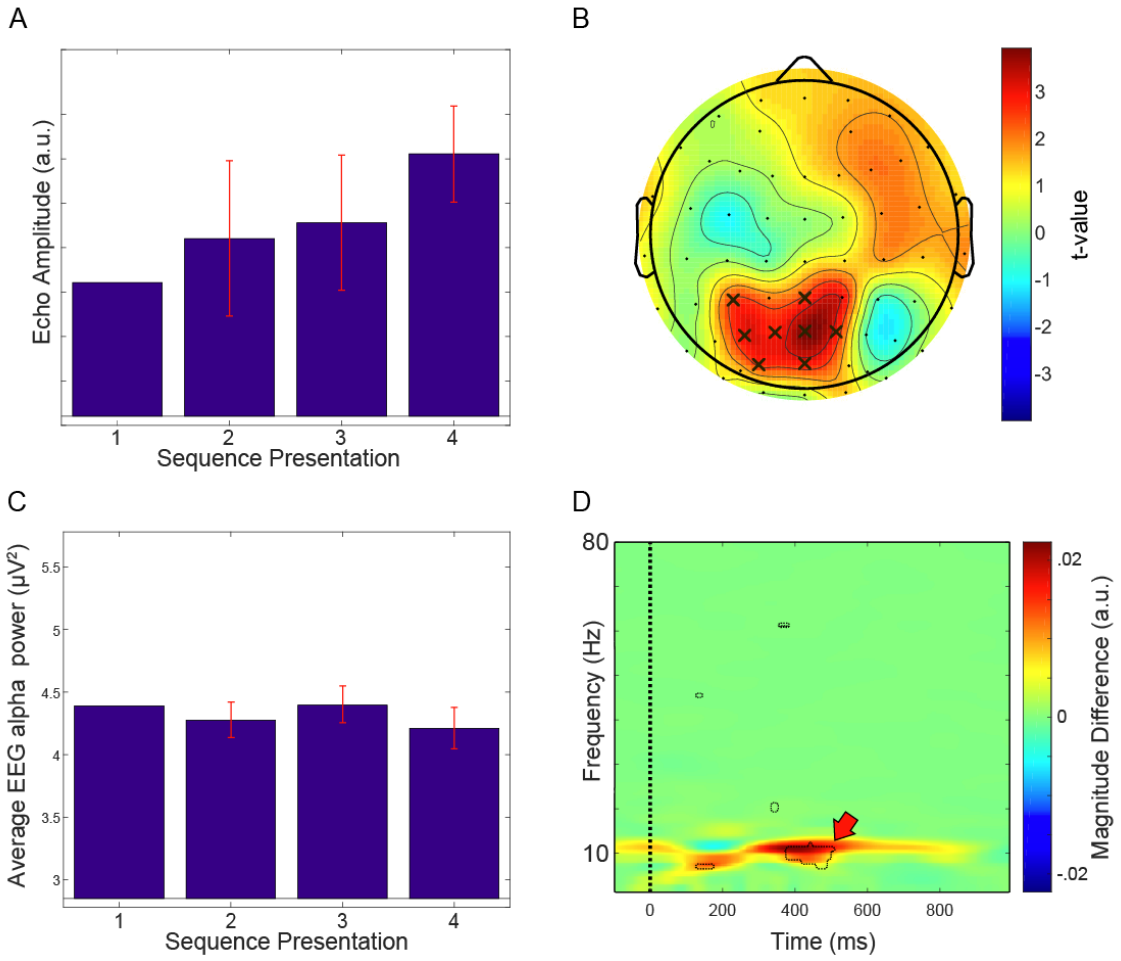

Figure 3-5: A. Experiment 1. Perceptual echo amplitude for each sequence presentation, averaged across all participants, for electrode POz. The echo amplitude displayed a linear increase across successive presentations. B. t-value map of the echo amplitude comparing presentation 4 to presentation 1 across all participants. The maximum difference in echo amplitude was located over occipital electrodes. (× indicates areas of significant difference between presentations, multiple t-test between all electrodes, Bonferroni corrected) C. Average of all participants' power from 8-12 Hz (μV2) for each sequence presentation for all trials for electrode POz. Average alpha power of the EEG time series was not affected by the repetition of a luminance sequence. D. Time-frequency analysis of the effect of sequence repetition on perceptual echo power. The plot shows the difference in time-frequency analysis of the perceptual echo between presentation 4 and presentation 1, averaged across luminance sequences and subjects, for POz. The echo differs

significantly between presentations 4 and 1 within the alpha frequency range from ~375 to 505ms (8-12Hz). The outlined area highlighted with a red arrow indicates significant differences using multiple t-tests, p<.05, False Discovery Rate (FDR) corrected. Average echo amplitude is measured in Arbitrary Units (a.u.); Average power is measured in microvolts squared. To effectively represent the variance within subjects, error bars show standard error of the difference between the first presentation and all subsequent presentations.

One might wonder whether the increase described above could be attributed to a general effect of the repetition of luminance sequences on the spectral power of the occipital EEG. To test this possibility, I computed the EEG amplitude spectra of the 3.125 epoch for every luminance sequence presentation and then averaged the amplitude spectra across each sequence presentation. I compared the average EEG amplitude spectra for each sequence presentation using polynomial contrasts, which did not reveal a significant linear relationship between presentation 1 to presentation 4 ($F(1, 13) = .65$, $p = .44$). Further post-hoc paired t-tests between the 4 presentations confirmed this result (all $ps > .28$). Thus, the increase in echo response amplitude with successive sequence presentations cannot be attributed to a general effect of EEG alpha response to these stimuli (see Figure 3-5c).

To examine the topography of the change in echo response from the first to the last presentation, I performed t-tests on echo amplitude between presentation 1 and presentation 4 across subjects and electrodes and plotted the t-values. The largest difference in echo response was observed over occipital electrodes (see Figure 3-5b).

I next computed the difference in echo response between presentations 1 and 4 in time-frequency space. This confirmed the repetition enhancement of the echo response, by showing that the alpha power of the echo response for presentation 4 was significantly larger than for presentation 1, with this effect concentrated within the period of the cross-correlation between

375 - 505 ms after sequence onset (ps < .05 FDR corrected, Figure 3-5d).

In the post-test questionnaire, all participants responded ‘no’ to all of the questions, ruling out explicit learning of a sequence as driving the increase in echo response with repeated sequence presentations.

Summarising, the results from Experiment 1 demonstrate that the amplitude of the perceptual echo response increases in a linear fashion with successive repetitions of a specific dynamic luminance sequence, even though participants were not aware of these repetitions. This finding shows that the visual system can encode temporal regularities defining repetitions of a specific luminance sequence. I next set out to determine the robustness and stability of this finding.

3.3 Experiment 2

Experiment 2 investigated the robustness of the increase in echo amplitude with repeated sequence presentations, across time and intervening sensory input. In this experiment, repeated (4) presentations of a specific luminance sequence were followed, firstly by an inverse luminance sequence, and secondly by another instance of the original (non-inverse) sequence. Each inverse sequence was created by inverting the luminance polarity, of the original sequence i.e. reversing black and white relative to the middle grey level (see Figure 3-6). This preserved the relative luminance values of the sequence while also maintaining a flat luminance power spectrum. I reasoned that, if the echo response reflects temporal regularity learning, the dissimilar luminance polarities of the inverse sequence should abolish any signature of sequence learning in the echo response, since the temporal contingencies of the sequence as a whole (luminance polarities) would be completely different while all other temporal and visual characteristics are preserved. I further reasoned that, if the learning process is robust across time and to intervening sensory input,

the echo response amplitude should recover or further increase in amplitude when the original sequence re-appeared following the inverse luminance sequence.

Experiment 2 was also designed to eliminate a potential confound in Experiment 1. In Experiment 1, each trial consisted of 4 presentations of the same sequence, and response sequences were randomly assigned to positions in-between experimental trials. This was done individually for each response sequence, resulting in the possibility of there being one, more than one, or zero, response sequences between any two experimental trials. This may have led to a potential, but highly unlikely, scenario that the distribution of response sequences fell by chance into a regular pattern between trials. In this unlikely situation participants may have developed strategies to deploy more attention towards the end of a trial in anticipation of a possible response sequence. Previous research has shown that increased attention can amplify the perceptual echo (VanRullen & Macdonald, 2012), raising the possibility that the increase in perceptual echo amplitude with repeated sequence presentations may have been caused by increased attention to the later presentations (i.e. presentation 4). To avoid this potential confound, in Experiment 2 the previous response sequences were no longer used. Instead, participants were required to estimate the average luminance level of the sequence in every trial, judging if the current sequence was on average brighter or darker than the fixation point. This modification meant that all sequences were 'response' sequences, ensuring that participants had to maintain equal attention to all sequences.

3.3.1 Methods

18 healthy students from the University of Sussex participated in this study (6 male, 20–36 years; mean age 25.2 years, normal or corrected-to-normal vision), none of which took part in Experiment 1. All provided informed consent before the experiment and received £10 or course credits as compensation. The experiment was approved by the University of Sussex ethics

committee. The apparatus was the same as in Experiment 1. There were a total of 120 trials. In the control condition (60 trials), a specific 3.125s sequence was presented for six times with 3s inter-sequence-interval. In the experimental condition (60 trials), a sequence was repeated four times. Then, on the fifth presentation, the inverse luminance sequence was presented. Following the inverse luminance sequence, the original sequence was presented again, making a total of six presentations of a sequence for each trial, see Figure 3-6.

In cases in which permutation t-tests did not display significant results I quantified how close to the null (no difference in amplitude) or alternative hypothesis (difference in amplitude) each result was by additional Bayes Factor (BF) analyses of paired sample t-tests using JASP (JASP Team, 2016) with a Cauchy prior of .707 half-width at half-maximum suggested by Rouder et al. (Rouder, Speckman, Sun, Morey, & Iverson, 2009).

Participants were asked to indicate whether the average luminance level of each sequence was brighter or darker than the luminance of the fixation circle, by pressing the left arrow for darker and right arrow for lighter, after each sequence presentation. The average luminance of a sequence was closely comparable across all sequence presentations (the standard deviations of the average luminance of all sequences relative to the entire luminance dynamic range of all sequences were .018% in Experiment 1 and .019% in Experiment 2). Therefore, participants could not use the average luminance of a sequence as a potential cue to encode a sequence. The luminance of the fixation circle was constant, meaning that the participant's judgements about the average luminance after each sequence were always based on similar information. Across all response sequences, participants rated 63.5% of sequences as being brighter than the fixation. One participant indicated for all sequences that the average luminance was brighter than the fixation, this may have been due to a perceptual bias when rating similar information across sequences, the data from this participant still showed a robust echo response and survived the exclusion criteria.

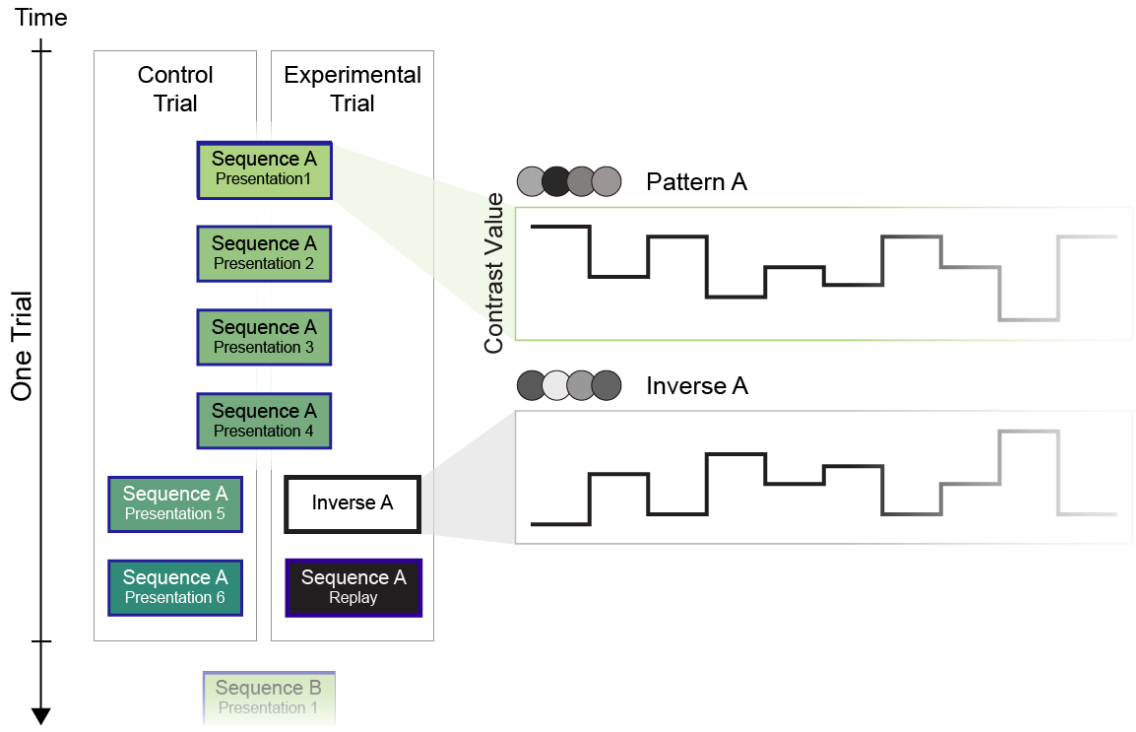

Figure 3-6: Experiment 2 trial structure. In a control trial, 6 presentations of the same sequence were shown. In an experimental trial, after 4 presentations of a given sequence, the inverse-polarity luminance sequence was presented, followed by a final replay of the original (non-inverse) sequence. Each sequence lasted 3.125s, separated by an inter-sequence-interval of 3s.

I adopted the same exclusion criterion as in Experiment 1, to exclude participants with no significant perceptual echo. Four participants were excluded using this procedure, meaning that 14 participants' data were retained for further analyses.

3.3.2 Results

To assess whether the main effect of sequence repetition on perceptual echo amplitude was present, I compared the amplitude of perceptual echo from presentation 1 to presentation 4 (data was pooled from experimental and control conditions) using a polynomial contrast analysis to test

the linear dependence between presentations 1 to 4. Although showing an unexpectedly high average echo amplitude for the 3rd presentation, the results remain supportive of a linear trend from presentation 1 to presentation 4, $F(1, 13) = 6.35$, $p = .026$, see Figure 3-7a.

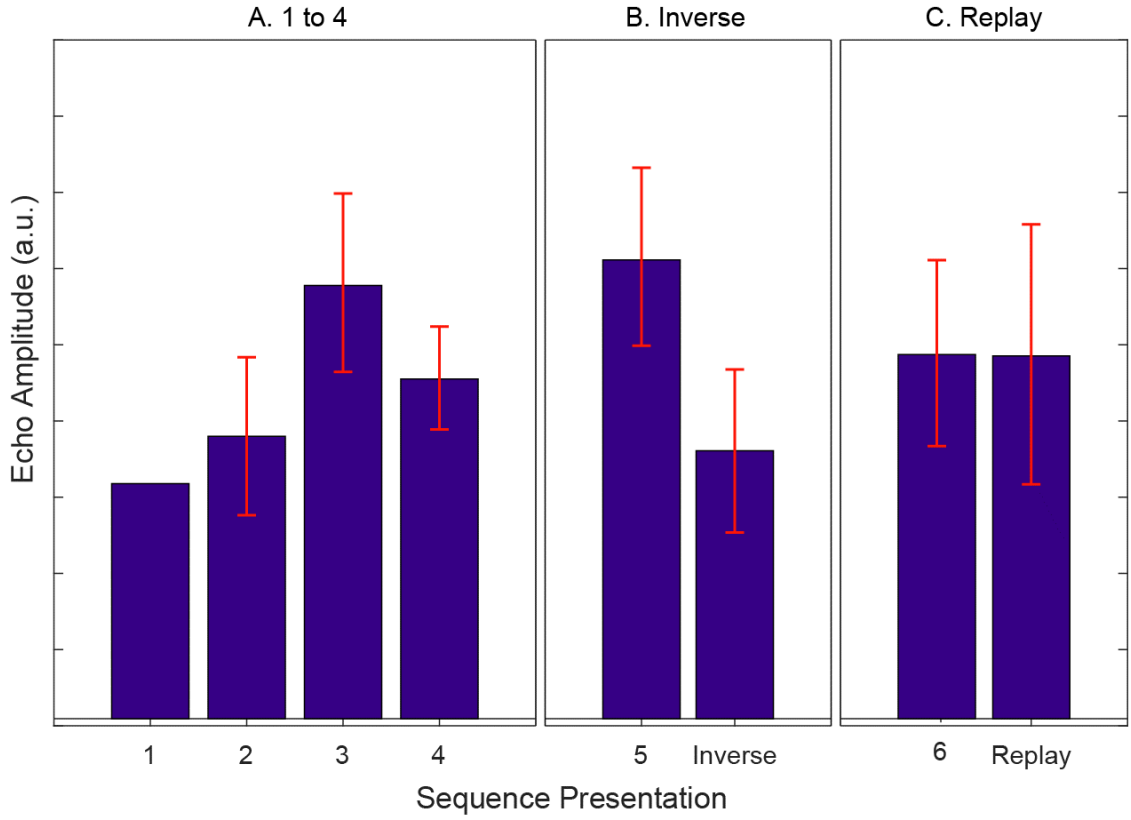

Figure 3-7: A. Echo amplitude as a function of repetition of luminance sequences. For the first 4 presentations of a given sequence, I observed a linear increase in echo amplitude with successive presentations, similar to Experiment 1 (data pooled across experimental and control trials). B. Critically, a subsequent 'inverse' sequence (experimental trials) showed a reduction in echo amplitude to a level similar to initial presentation of a luminance sequence, as compared to a fifth presentation of a non-inverse sequence. C. Re-presenting the original sequence following the inverse sequence (Replay, experimental trials), showed a recovery of echo amplitude indistinguishable from a 6th successive presentation. Perceptual echo amplitude is measured in Arbitrary Units, a.u. To effectively represent the variance within subjects, error bars show standard error of the difference between the first presentation and all subsequent presentations.

I next examined whether presentation of an inverse luminance sequence would abolish the

increase in echo amplitude seen after four successive presentations. I compared the echo amplitude elicited by the 5th presentation between experimental (i.e. the inverse luminance sequence) and control conditions. In this comparison, the two echo amplitudes were both evoked by the fifth presentations: either an inverse sequence (experimental condition), or another presentation of the same (non-inverse) sequence (control condition). If the new sequence is able to interrupt the repetition effect, we should expect the analysis yields a null-result. The echo amplitude elicited by the inverse sequence was significantly lower than elicited by a 5th (non-inverse) presentation (two-tailed paired t-test, $t(13) = 4.36$, $p < .002$, bootstrap, 10,000 resamples)(see Figure 3-7b). I also compared the echo amplitude elicited by the inverse sequence and presentation 1 in experimental trials and found no significant difference in echo amplitude between the two presentations (two-tailed paired t-test, $t(13) = .02$, $p = .71$, bootstrap, 10,000 resamples). Traditional statistics hypothesis testing（i.e., t-test) does not provide a quantitative measure about how strongly the data supports the null hypothesis. I additionally used Bayes factor analysis to further evaluate to what extent the echo amplitude elicited by the inverse sequence supported the null hypothesis (i.e. a conclusion of no difference in echo amplitude between the 1st presentation and inverse sequence.) or the alternative hypothesis, (i.e. echo amplitude was different between the 1st presentation and inverse sequence). As expected, I found BF = .29 (less than .33), which moderately supports the null (i.e., no difference) over the alternative hypothesis (Dienes, 2011).

Together these results indicate that an inverse luminance sequence, carefully controlled for a range of perceptual and temporal properties, was processed in a similar manner as a new luminance sequence, in terms of perceptual echo. This confirms that the visual system is encoding precise sequence information, rather than only general temporal properties of luminance sequences (e.g. time-frequency dynamic, auto-correlation) or its visual characteristics (e.g. luminance range and

variance), since these more general features are shared with the inverse luminance sequence.

I next compared the topography of the echo response amplitude for the 5th sequence presentations between experimental and control trials. I performed t-tests on echo amplitude between these presentations across subjects and electrodes and plotted the resulting t-values, see Figure 3-8a. Similar to Experiment 1, the largest difference in the echo response was found over occipital electrodes, with a maximum over POz.

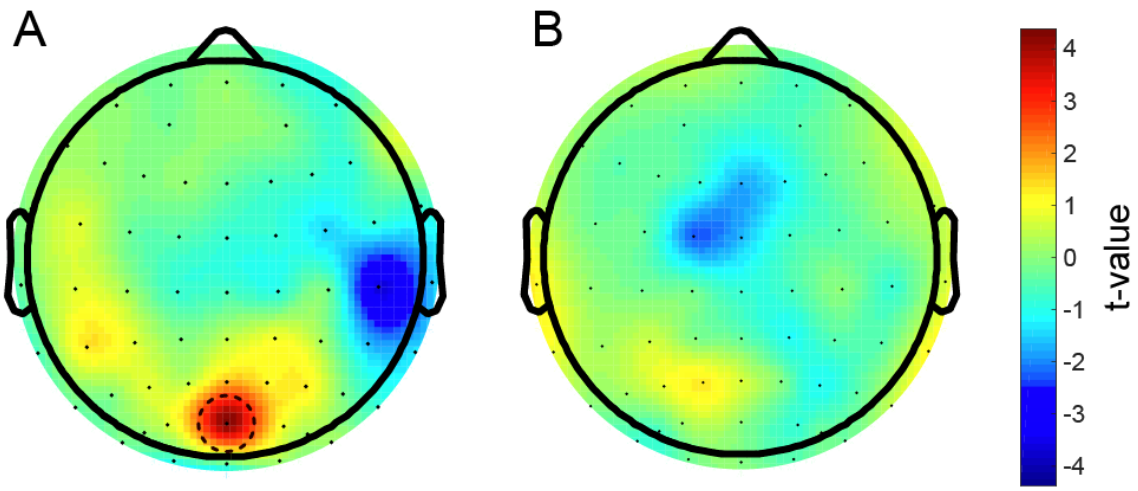

Figure 3-8: A. t-value map of echo amplitude difference between presentation 5 (control condition) and the inverse luminance sequence (experimental condition) for all subjects, in Experiment 2. A significant difference was centred over occipital electrode POz (outlined, FDR corrected). B. t-value map of echo amplitude difference between presentation 6 (control condition) and the replay sequence (experimental condition) for all subjects, in Experiment 2. There were no significant differences in echo amplitude between the two presentations.

I then examined whether sequence information encoded across presentations 1 to 4 could persist even after the presentation of an inverse luminance sequence. I reasoned that if such information does persist, re-presenting the original sequence after the inverse luminance sequence should restore the amplitude of the perceptual echo to a level comparable to six sequential presentations of a particular sequence. I therefore compared the echo amplitude elicited by the

'replay' sequence in the experimental condition, and the 6th presentation in the control condition. No significant difference was observed (Figure 3-7c); indeed, echo amplitudes for these two presentations were almost identical (t = .01, p = .99). I computed BF to evaluate whether the echo amplitude elicited by the replay sequence was supportive of the null hypothesis (i.e. no difference in echo amplitude between the replay and presentation 6) or the alternative hypothesis, (i.e. echo amplitude is different between replay and presentation 6). I expected that the BF should support null-hypothesis more than alternative hypothesis. A BF = .27 provides strong evidence (less than .33) for accepting the null over the alternative hypothesis, indicating that sequence-specific information about a particular luminance sequence persists, for over 9 seconds, even in the presence of intervening visual input.

Finally, I examined the topography of the difference in echo response amplitude between presentation 6 (control) and replay (experimental) sequences. I performed t-tests on echo amplitude between these presentations across participants and electrodes and plotted the resulting t-values, see Figure 3-8b. There were no significant differences at any sensor, indicating that the echo response to the replay luminance sequence displayed a similar topography and magnitude as presentation 6 (All BF < .79, supporting the null hypothesis).

3.4 Discussion

In two experiments, I investigated the functional relevance of the perceptual echo response, testing the hypothesis that it reflects a predictive processing mechanism which can encode and learn dynamic visual sequences within the visual cortex.

Supporting this hypothesis, Experiment 1 showed that the perceptual echo response is enhanced by repetitions of an identical rapid luminance sequence, suggesting that the perceptual echo reflects a

neural signature of temporal regularity learning.

Control analyses using shuffled data excluded the possibility that increases in perceptual echo amplitude could be attributed to general changes in induced alpha-band EEG responses resulting from sequence repetitions.

Experiment 2 established that the repetition-dependent changes in perceptual echo reflect specific sequence information. Following 4 presentations of a specific sequence I presented an inverse luminance sequence, which preserved all non-sequential spectral and temporal properties of the original sequence. Strikingly, the echo amplitude for the inverse sequence returned to a level comparable to presentation of a novel sequence (see Figure 3-7b). When the original stimulus sequence was presented again (following the inverse sequence), the echo amplitude recovered to a level consistent with the number of presentations of this sequence, demonstrating that information about an encoded sequence persisted for over 9 seconds and was robust to intervening visual input.

Behavioural studies have demonstrated that humans are able to learn temporal sequences presented at different rates (1.7-8.3 Hz) consisting of spatiotemporal information (Song, Jr, & Howard, 2008), combinations of visual features (colour and spatial) (Gheysen, Van Opstal, Roggeman, Van Waelvelde, & Fias, 2011) and object orientations (Luft, Meeson, Welchman, & Kourtzi, 2015). These results extend these findings by showing that the visual system is capable of sequence learning even with stimuli presented orders of magnitude faster (160 Hz) than previously used. This sensitivity to rapidly changing input is in line with known properties of the human auditory system, where auditory sequence learning has been described behaviourally for random and meaningless input signals (Gaussian random noise). Notably, as in the present study, signatures of sequence-specific learning were found to persist for seconds and be robust to intervening auditory inputs (Agus et al.,(2010).

Regarding neural substrates of sequence learning, both human and animal studies have implicated brain regions outside primary sensory areas. For example, in monkeys, information about spatiotemporal sequences is encoded in inferior temporal cortex and also V4 (Li & DiCarlo, 2008; Meyer & Olson, 2011). However, Gavornik and Bear (2014) found that, in mice, it was possible to detect the encoding of spatiotemporal sequence information early in the visual stream, within the primary visual cortex. These findings are compatible with the localisation of the perceptual echo – and its increase across repeated sequence presentations – to (human) visual regions.

What mechanisms could be responsible for the increase in echo response that I observe with successive presentations of luminance sequences? One possibility is that this increase may be a consequence of the increasing similarity across sequence repetitions between luminance changes in the stimulus sequence and changes in the evoked EEG response. This would result in stronger cross-correlations between the EEG and the luminance sequence, with successive presentations, and hence an increase in echo amplitude. This could be thought of as 'sharpening' of the neuronal assemblies representing visual sequences, reflecting increasingly precise representations of repeated sequences. Supporting this interpretation, previous studies on perceptual learning have shown that repeated exposure to the same visual input sharpens the responsiveness of neuronal assemblies representing these inputs (Seitz & Dinse, 2007). For instance, monkeys trained on an orientation discrimination task show a post-training decrease in response variability and an increase in the slope of orientation-tuning curves in V4 for the trained orientations (Yang & Maunsell, 2004). Similarly, the phenomenon of repetition suppression may be caused by a sharpening or tuning mechanism, which occurs when repeated exposure to a stimulus leads to a more precise and more efficient neural representation of that stimulus (Desimone, 1996; Kok et al., 2012). At a mechanistic level, neuronal sharpening is thought to depend on Hebbian processes,

among which spike-timing dependent plasticity (STDP) is particularly significant when dealing with temporal relationships (Bi & Poo, 1998). Markram et al., (1997) found that STDP displayed the highest speed and efficacy when spike timing was reliable. Additionally, STDP has been shown to make neurons extremely sensitive to repeating spatio-temporal patterns (Guyonneau, VanRullen, & Thorpe, 2005; Markram, Lübke, Frotscher, & Sakmann, 1997; Masquelier, Guyonneau, & Thorpe, 2008a, 2008b). The repeated presentation of a luminance sequence may have increased both the sensitivity and reliability of neural responses facilitating the rapid encoding of sequence information found in this study.

An alternative explanation for the primary finding is that the *amplitude* of specific evoked EEG responses to the luminance sequence could increase across successive repetitions. This would lead to an increased signal to noise ratio and thus also to stronger cross-correlations between the EEG time series and luminance sequences. While the opposite phenomenon is commonly observed (repetition suppression), a wide range of studies investigating perceptual learning have found that the neural response to a stimulus can also be enhanced by repeated exposure to identical sensory input (Karni & Sagi, 1991; Vogels, 2010). However, the results do not support this alternative interpretation since I found no difference in alpha power in the raw EEG signal across repetitions of the same luminance sequence, as would be expected if the amplitude of the alpha-band evoked EEG signal was driving the increases in echo response across sequence repetitions (Figure 3-5c).

In summary, the enhancement of the echo response with successive presentations of a sequence is most plausibly due to an increase in similarity between luminance changes in the stimulus sequence and changes in the evoked EEG response. A candidate mechanism for this process is STDP, which has the effect of sharpening the population response for each specific sequence, leading to an enhanced echo response with successive presentations of a sequence.

Could the perceptual echo be a signature of active visual processing? A striking feature of the

echo response is its periodicity, this feature suggests that the visual system is actively engaging in repeated patterns of activity in response to ongoing sensory stimulation. One might speculate that periodicity of the perceptual echo reflects an underlying iterative learning process that updates and replays temporal representations at a rate defined by occipital alpha. This interpretation is motivated by the influential 'predictive processing' account of perception and brain function, in which the brain continuously generates and updates predictions about incoming sensory signals (Clark, 2012; Karl Friston, 2005; Hohwy, 2013; Seth, 2014). Extended formulations of this account suggest that hierarchical predictive generative models update prior knowledge by extracting and encoding hidden spatial and temporal regularities in the environment (Tenenbaum et al., 2011). In this study, increases in echo response amplitude with successive repetitions of a sequence demonstrate that participants were implicitly learning information about each sequence, which may reflect perceptual predictions (Bayesian priors) being updated with each sequence. In this view the echo response may reflect an iterative process that updates priors about the luminance dynamics of a sequence, communicating perceptual predictions at rate defined by the alpha frequency band. Further research will be needed to tie perceptual echo to predictive processing more substantively, for example by manipulating expectations about temporal regularities, within or across sequence presentations.

Another question that was not fully explored in the present study was the persistence of the encoded sequence information. These results demonstrated that this information can persist for more than 9 seconds, however the constraints on retention time of this information was not systematically examined. If as I suggest the repetition of a luminance sequence selectively strengthens synaptic pathways through STDP, then I expect that a long-term enhancement of the echo response should be detectable. I also predict that it should be possible in animal studies to disable the sequence learning effects found in the present study by using optogenetics to 'switch'

off receptors critical for long-term potentiation (LTP) , such as NMDA receptors. In such studies there should be no effect of sequence repetitions on the perceptual echo.

To further investigate the perceptual echo responses role in regularity learning within a predictive processing framework, future studies could investigate how manipulating the predictability of a sequence affects the echo response. Another possibility to verify the perceptual echo responses involvement in predictive processing would be to test whether it plays a role in periodically transmitting prior information within an alpha defined frequency. By using transcranial alternating current stimulation (tACS) over the visual cortex it is possible to entrain a specific alpha phase and frequency. This manipulation would allow for the time locking of sequence onset to a specific alpha phase. I predict that, if the alpha cycle is critical for the transmission of prior information, then sequences presented in-phase with alpha should lead to an enhanced echo response in later repetitions of the same sequence.

Future studies could also investigate a more diverse range of stimuli in the context of perceptual echo, since the current study, used only a low-level visual feature: luminance. By exploring sensory dimensions including orientation, colour spatial information, it will be possible to examine the extent to which perceptual echo responses (and their enhancement with sequence repetitions) generalise. These results also motivate investigation into the persistence of the encoded sequence information. These data show that this information can persist for more than 9 seconds, however the constraints on retention time of this information were not systematically examined. If as I suggest the repetition of a luminance sequence selectively strengthens synaptic pathways through STDP, a long-term enhancement of the echo response may be expected.

3.5 Modelling the perceptual echo using a predictive processing framework

3.5.1 Introduction

The previous section demonstrated that perceptual echo was modulated by repetitions of a specific luminance sequence, suggesting that it reflects a signature of regularity learning within the visual system. This result supports predictive processing accounts of the low-level visual system, which suggests that perception is an active learning and inference process. In this section, I propose a simple predictive coding model that captures several aspects of the perceptual echo and demonstrates that the perceptual echo can be generated by this simple system using minimal assumptions.

3.5.2 The aims of the model

This model intends to simulate two intriguing characteristics of perceptual echo. First, the 10Hz periodical fluctuation in the cross-correlation between luminance sequences and EEG signals only shows at positive lags (after stimulus onset) but is absent at negative lags (see Figure 3-2). Cross-correlation theorem states that the power spectrum of the cross-correlation of two time series is equal to the product of the individual power spectrum, where one of them has been complex conjugated (Bracewell, 2004). Therefore, one may expect that the perceptual echo may be driven by the alpha oscillations in the raw EEG. However, the perceptual echo is only visible after stimulus onset at positive lags. In terms of the cross-correlation analysis this feature indicates that all of the sequence information is maintained and embedded in the EEG signal, leading the peak in the cross-correlation. This observation implies that the echo reflects a ‘memory trace’ in the human

visual system, rather than being driven by the raw EEG alpha, which would affect the cross-correlation at both positive and negative lags.

This result is also consistent with the experimental finding of this Chapter that repetitions of stimulus sequences only enhanced the echo response (cross-correlation) and was not observed in changes in the alpha power of the raw EEG, suggesting that the raw EEG alpha is not related to the regularity learning process reflected by the perceptual echo. Therefore, a crucial aim of the model is to simulate the asymmetrical cross-correlation pattern between positive and negative lags.

The second aim of this model is to account for the results of a pilot experiment, in which I found a different temporal dynamic of the perceptual echo when the stimulus sequence was presented centrally instead of peripherally. In this pilot study, the experiment procedures and design were the same as Experiment 1 and Experiment 2 as described in Chapter 3.2 and 3.3 except that the luminance sequence was presented at the centre of the visual field instead of peripheral areas. Data from 5 subjects showed that when the stimulus sequence was presented centrally the morphology of the echo response changed dramatically, showing a more abrupt and greater deflection at early lags followed by a quick decay at later lags (Figure 3-9) compared to the echo responses elicited by peripheral stimulation (see Figure 3-2).

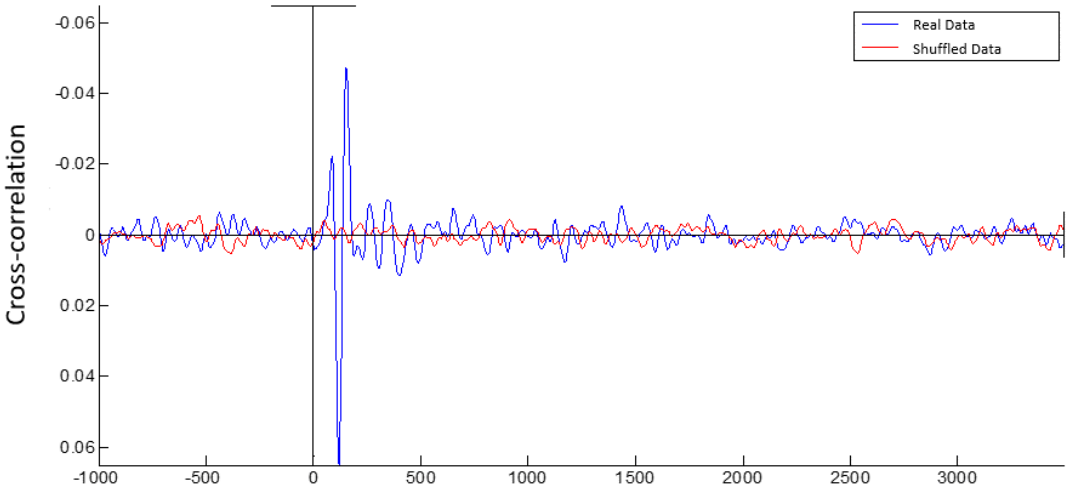

Figure 3-9 Perceptual echo elicited by stimulus sequences presented in the centre of the visual field, data from a pilot study. Perceptual echo showed a sharp fluctuation at early lags (before 200ms) and a quick decay.

Examining these two characteristics, the first that the perceptual echo displays multiple peaks at positive lags, suggests that the sensory sampling of the external luminance values persists in the visual system and influences subsequent EEG signals around every 100ms (10Hz cycle). Therefore, I assume that EEG signals are influenced by both the current luminance value of the sample as well as the memories of previous luminance samples (this can also be viewed as a Bayesian prior). The second characteristic, that the eccentricity of stimulus sequences influences the perceptual echo response suggests that the eccentricity influences the weighting between the current visual luminance sample and the memories of previous luminance samples on the EEG signal. It is well known that the spatial resolution (precision) of central and peripheral vision differ greatly (Loschky, McConkie, Yang, & Miller, 2005). Therefore, the sampling precision may plausibly be a factor in influencing how the visual system combines the current visual sample and memories of luminance values, which may be best explained by a Bayesian/predictive processing framework.

I propose a simple predictive coding model that uses minimal assumptions in an attempt to account for the two perceptual echo characteristics outlined above. An import feature of Bayesian

inference and the predictive processing framework is that information is integrated based on the weighting derived from their precision (Hohwy, 2012; Knill & Pouget, 2004). The more reliable information source is weighted more for computing the result of the integration (Bayesian posterior, also see Chapter 1). Therefore, central vision, which has a high precision may lead to a stronger weighting on the current sensory sample compared to the prior. When the precision is high, the perceptual echo should display higher amplitudes initially and then decay rapidly, due to the influence from priors being low, indeed I found the result when using stimuli presented centrally in a pilot study. In contrast, peripheral vision, which has lower precision may rely more on prior than new sensory evidence. When the precision is low, the perceptual echo should be longer lived as the memory of a sample is able to impact perceptual inference on more following samples.

3.5.3 Model computation

This model is inspired by Kalman filter (Kalman, 1960), a prediction-based perceptual inference system. The model consists of two units, a prediction unit and a prediction error unit (Figure 3-10a). The prediction unit generates predictions about the next sensory sample, in this case of the next luminance value and sends the value to the prediction error unit via a top-down pathway. After about 50ms, information transmission time, the prediction error unit samples the next external luminance value and computes the difference (prediction error) between the top-down luminance prediction about the current luminance value and the actual luminance value from the current sensory sample. The prediction error value is then sent to the prediction unit through a bottom-up pathway with about 50ms transmission time. The prediction unit integrates and updates the prediction by computing a weighted average between the prediction error and the current prediction using a weighting factor w. Finally, the new prediction is send to the prediction error unit

to initiate the next iterations of the cycle (Figure 3-10b). The model assumes that the information transmission time is 50ms and, therefore, a cycle of 100ms is created.

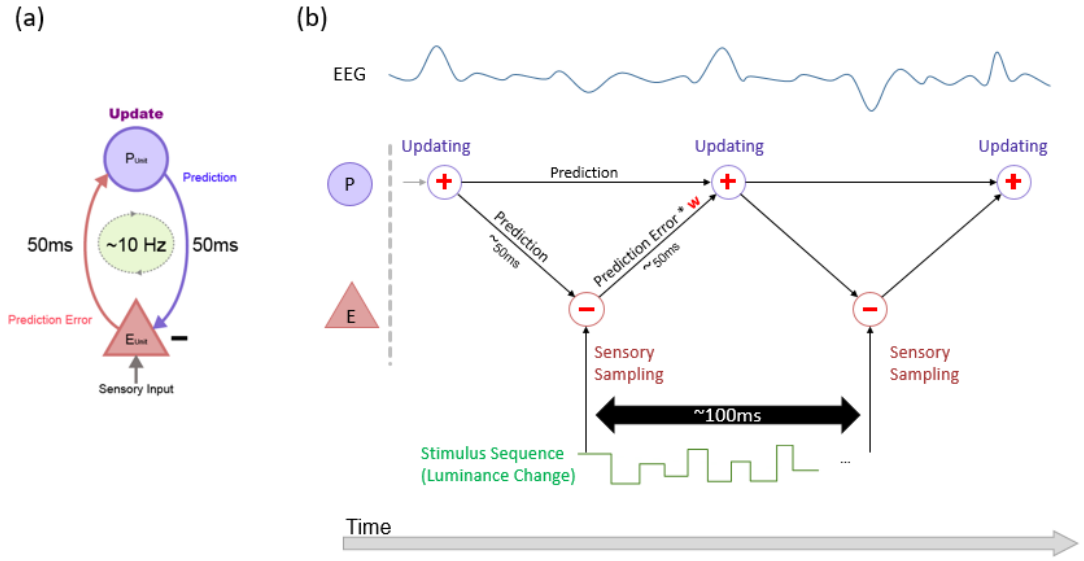

Figure 3-10. A predictive coding model of perceptual echo (a) The model consists of a prediction unit and a prediction error unit. The prediction unit sends prediction though a top-down pathway. The prediction error unit sends a prediction error signal via a bottom-up pathway. The model assumes that the information transmission time is 50ms and, therefore, a cycle of 100ms is created. (b) Illustration of the model processing steps across time. The model samples external luminance values every 100ms. At every sensory sampling point, the low level prediction error unit computes a prediction error by computing the difference between a top-down prediction and the current sensory input. The high level unit then updates the predictions by computing the weighted average between the current prediction and the prediction error using a weighting factor w. This updating and prediction error computations causes signal changes in the raw EEG.

3.5.4 Model implementation and simulation

Luminance sequences were generated using the same parameters and methods as the two EEG experiments of this chapter. Therefore, each luminance sequence was 3.125s and the range of luminance values was between 0 to 255.

The EEG time series in each trial was simulated using Eq. 3-1.

$$EEG\,(t) = \begin{cases} seq(t) - EEG(t_{update}^{i-1}) + \epsilon_{EEG} & if \quad t \in t_{sample} \\ EEG(t_{update}^{i}) + w \times EEG(t_{sample}^{i}) + \epsilon_{EEG} & if \quad t \in t_{update} \\ b + \epsilon_{EEG} & otherwise \end{cases} \quad (3\text{-}1)$$

$$\epsilon_{EEG} \sim N(\mu_{EEG}, \sigma_{EEG}) \quad (3\text{-}2)$$

EEG fluctuation was determined by an updated prediction and a prediction error at updating (t_{update}) and sampling (t_{sample}) time, respectively. Otherwise, EEG fluctuation was only influenced by Gaussian random noise (see Eq. 3-1 and 3-2) added to a baseline EEG activity b. ϵ_{EEG} is a Gaussian random noise sampling from a Gaussian distribution with a mean μ_{EEG} and a standard deviation σ_{EEG} added to each EEG time series data point (Eq. 3-2). In this simulation, μ_{EEG} was fixed to 0 and σ_{EEG} was fixed to 16% of the dynamic range of luminance values. The baseline EEG activity was fixed to the midpoint of the luminance dynamic range (i.e. 127.5).

When the predictive processing model sampled a luminance value at time t (i.e., $t_{sample}^{i} \in t_{sample}$), a prediction error was computed by subtracting the last prediction $EEG(t_{update}^{i-1})$ from the sampled luminance value $seq(t)$ and reflected in a EEG amplitude change (Eq 3-1). The prediction error was then used to update a new prediction at the next update time t_{update}^{i} after information transmission time $t_{transmission}$ (Eq. 3-3).

$$t_{update}^{i} = t_{sample}^{i} + t_{transmission} \quad (3\text{-}3)$$

$$t_{sample}^{i+1} = t_{update}^{i} + t_{transmission} \quad (3\text{-}4)$$

$$t_{transmission} \sim N(\mu_{transmission}, \sigma_{transmission}\ ) \quad (3\text{-}5)$$

As mentioned above, I assumed the after information transmission time between the prediction and the prediction error units is about 50ms. Therefore, I sampled $t_{transmission}$ from a

Gaussian distribution with a mean $\mu_{transmission} = 50ms$ and a standard deviation $\sigma_{transmission} = 10ms$ (Eq. 3-5).

At updating time t^i_{update}, a updated prediction was computed by summing the current prediction $EEG(t^i_{update})$ and a weighted prediction error $w \times EEG(t^i_{sample})$ computed from the last sample time t^i_{sample} and reflected in a EEG amplitude change (Eq. 3-1). The updated prediction was then used to compute next prediction error after information transmission time (Eq. 3-4).

In each trial, the initial luminance prediction was set to 0 because no stimulus was presented before the sequence onset (dark background). The model started to sample at t = 0ms, i.e., $t^1_{sample} = 0ms$.

To simulate the temporal dependence between adjacent sample points found in real EEG signals, a 1D Gaussian filter with a 31ms window size and a 19ms standard deviation was applied to the simulated EEG time series in each trial.

I tested the weighting (w) parameter value changing it from 0.1 (sensory evidence is not reliable) to 0.9 (sensory evidence is very reliable) to simulate different reliabilities of sensory evidence.

The simulation was run for 1000 trials for each w. Cross-correlations were computed and averaged across trials using the same methods as the two EEG experiments.

3.5.5 Critical model assumptions

The model makes two major assumptions. First, based on the fact that perceptual echo mainly fluctuates at about 10Hz, the model assumes that information transition time between prediction and prediction error units through top-down and bottom-up pathways takes about 50ms, and therefore, takes 100ms to complete a cycle for each iteration. Second, the notion that the human

visual system samples visual information in a discrete manner at about a 10Hz frequency has been proposed and empirical evidence supporting this notion has been accumulating (VanRullen & Koch, 2003). According to this idea, the model assumes the visual sampling is discrete. Sensory sampling only occurs when the prediction error unit needs to compute a prediction error. Based on the first assumption, the interval between visual samples is about 100ms. Therefore, this assumption yields the visual sampling frequency of this system is 10Hz.

3.5.6 Results

Figure 3-11 shows the results of this model. As can be seen this simple predictive processing model captures two important aspects of echo responses from the experimental studies of this chapter.

Firstly, the periodical cross-correlation fluctuation only was found in the positive time lags. This result confirms that perceptual echo is not only caused by the alpha oscillation in raw EEG data but by stimulus-relevant memory in a perceptual system.

Second, the precision of sensory input modulates the temporal dynamics of the echo response (Figure 3-12). As can be seen from this figure higher sensory weights elicited an echo response with greater amplitude at early time lags followed by a rapid decay. In contrast, lower sensory weights elicited smaller echo responses at early time lags but sustained a longer response. To quantify the decay rates, I fitted a two-parameter exponential function ae^{bx} to the envelope of each cross-correlation from the pick value to the end (Figure 3-13). The exponential terms b (growth rate) from the fitting result showed a nearly monotonic increase along w decrease, suggesting a faster decay when w is high than when w is low. This result fits the empirical data showing that the eccentricity of a stimulus sequence modulates the temporal dynamics of the echo response.

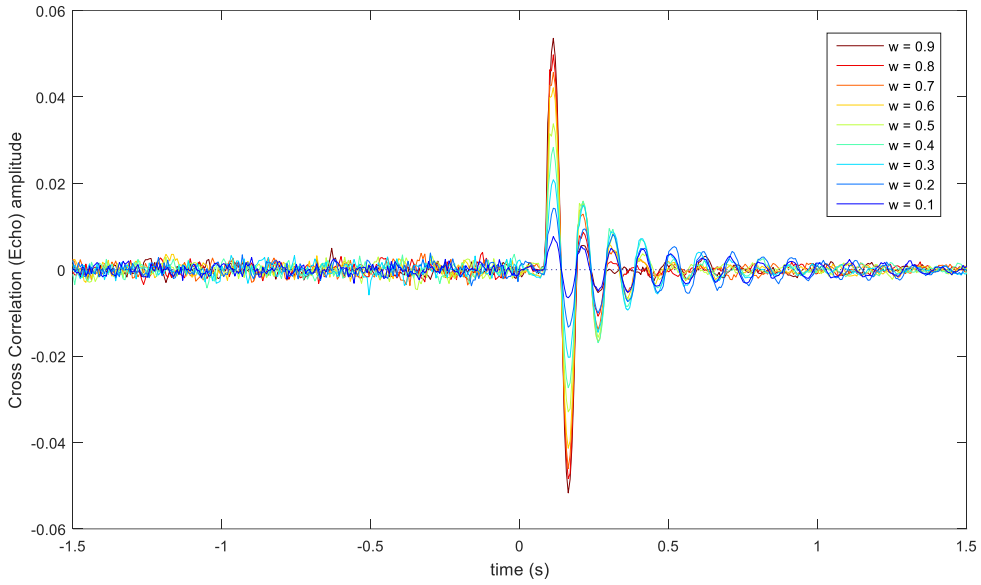

Figure 3-11. The cross-correlation between simulated EEG and repetitions of a luminance sequence averaged across trials (for details of trial number etc. refer to methods of Experimental sections). The parameter w represents the weights of sensory evidence (precision) in the model, with higher values representing higher precision of sensory input. Abscissa represents cross-correlation lags in time.

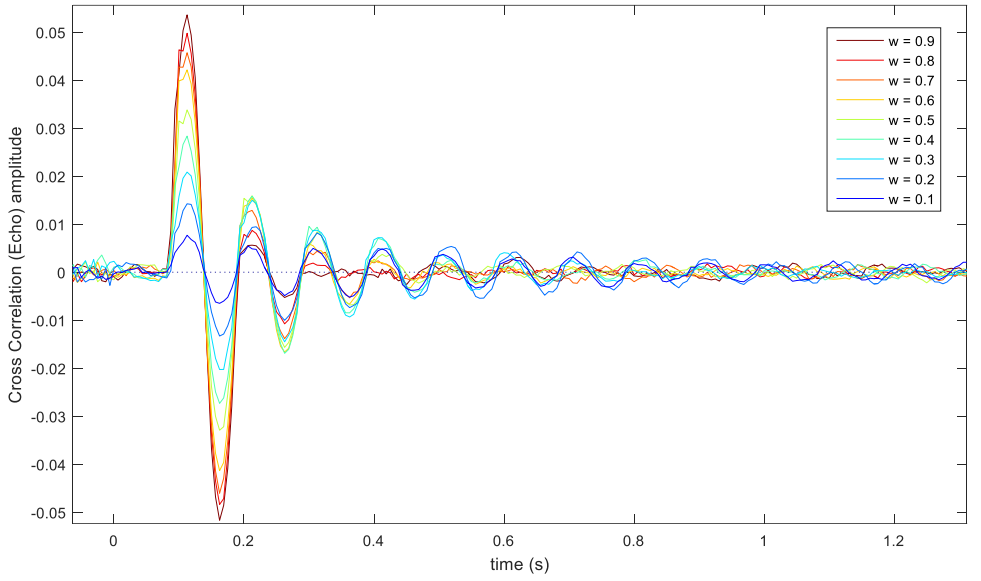

Figure 3-12 A expanded view of Figure 3-11 within positive time lags.

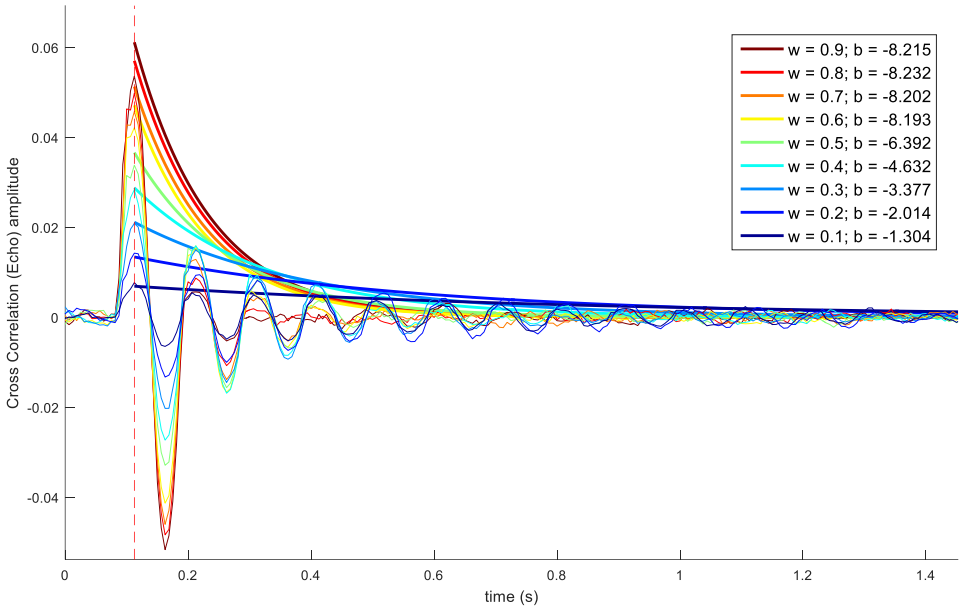

Figure 3-13 Fitting envelopes of cross-correlation using exponential functions. Cross-correlation data were selected from the maximal value (red dot line) to the end of each cross-correlation series. The exponential terms b (growth rate) from the fitting result showed a nearly monotonic increase along w decrease.

To further quantitatively test the difference between perceptual echo at positive and negative lags, shuffled cross-correlation data were recruited by computing cross-correlation between EEG time series and luminance sequence from different trials. Similar to the experimental results showed in Section 3.2, shuffled cross-correlation data showed no perceptual echo response (Figure 3-14), indicating the sequence-specific nature of perceptual echo. More importantly, even though 50ms transmission time was explicitly assigned to the model which may cause a large alpha oscillation in the simulated EEG data, shuffled cross-correlation data with no phase coherence between trials (Figure 3-2) was still unable to generate perceptual echo.

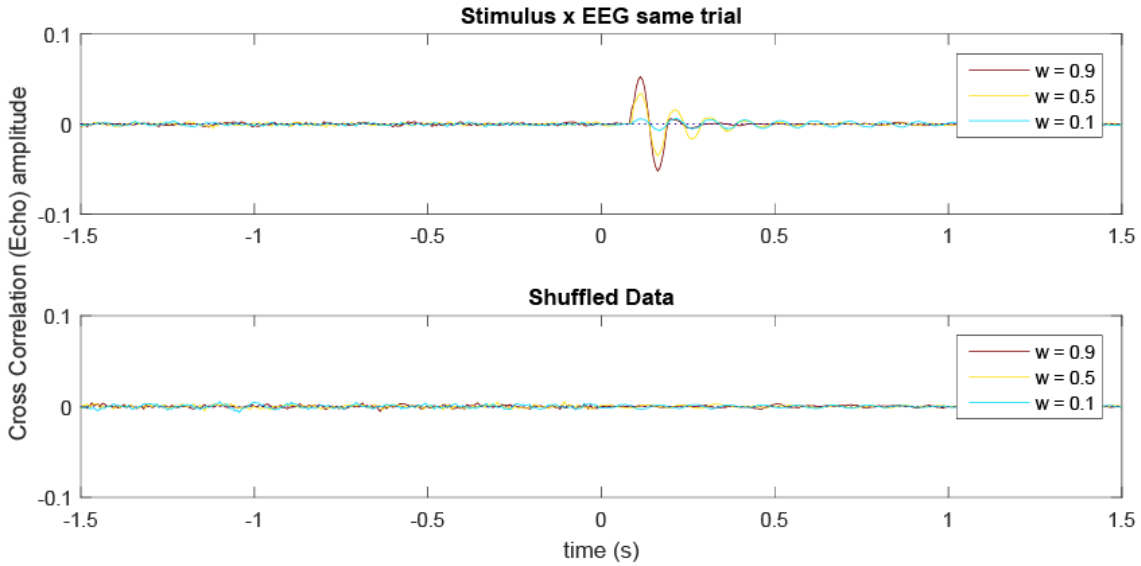

Figure 3-14 Comparison between actual data (cross-correlation between a EEG time series and a luminance sequence was computed from the same trial) and shuffled data (from different trials). Shuffled cross-correlation data showed no echo-like periodic oscillation.

Similar to the examination of perceptual echo significance for each participant in Section 3.2, I compared cross-correlation data distributions at positive and negative lags ranging from -0.7s to 0.7s for actual and shuffled data with w = 0.1, 0.5 and 0.9 (Figure 3-15). The result showed that only non-shuffled data with w = 0.1 and 0.5 showed a significant difference of cross-correlation data distribution between positive and negative lags ($p < 10^{-5}$ and $p < 10^{-3}$, respectively), suggesting perceptual echo only presented at positive lags. The null result of actual data when w was 0.9 confirmed that with large sensory weight perceptual echo only presented in a short period and decayed very quickly. In sum, the model successfully simulated the experimental phenomenon that perceptual echo only was found in the positive time lags, suggesting stimulus-relevant memory in a perceptual system.

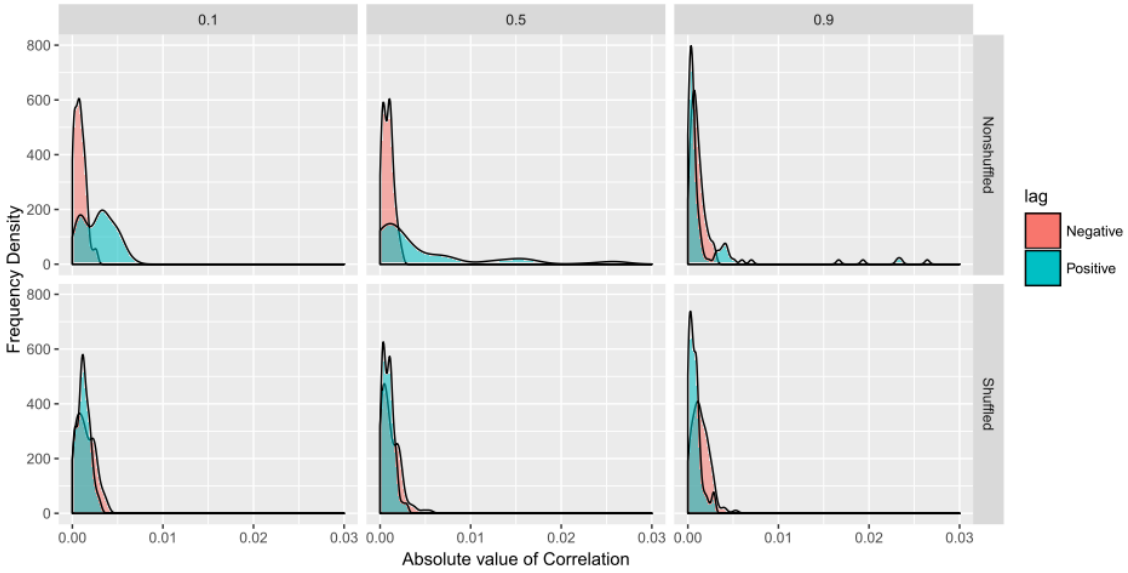

Figure 3-15 Echo response data distribution from non-shuffled (nonshuffled, the first row) and shuffled (the second row) data. Three different weighting factors (w = 0.1, 0.5, and 0.9, indicated by columns) were used. When w is 0.1 and 0.5, cross-correlation data distributions from positive lags were significantly different from data distributions from negative lags.

3.5.7 Discussion

This simple model provides a computational account of how a predictive coding system generates the perceptual echo. As can be seen from this section the resemblance between the echo produced by this model and those found experimentally is striking, suggesting that the predictive processing interpretation of the perceptual echo as an iterative learning process that updates and replays temporal representations is well founded. This model also provides an explanation for the basic characteristics of the perceptual echo, via the change in weighting between a prediction and a prediction error unit. Finally, this model has explained two essential characteristics of perceptual echo that is best accounted for by the predictive processing framework.

3.6 Conclusion

The perceptual echo is a long-lasting periodic reverberation in the EEG response to dynamic

visual stimulation, reflected by an alpha-band peak in the cross-correlation function between a rapidly changing random dynamic luminance sequence and the raw occipital EEG response. I investigated the functional properties of the perceptual echo, finding that it can be enhanced by repeatedly presenting the same visual sequence, indicating that the human visual system can rapidly and automatically learn temporal regularities embedded within such fast-changing dynamic stimulus sequences. By comparing echo responses for inverse and non-inverse luminance sequences, I further showed that the increase in echo response was sensitive to specific sequence information. Finally, I show that the encoded sequence information can persist over many seconds even in the presence of additional intervening sensory input. Using a simple predictive coding model, with minimal assumptions, I show that the model can capture many of the features of the perceptual echo, further supporting the notion that predictive processing underlies the perceptual echo.

Together, these results provide strong evidence for the existence of a previously undiscovered predictive temporal regularity learning mechanism that appears to be governed by predictive processing, this process being reflected by the perceptual echo.

Chapter 4
Examine temporal prediction from duration processing

4.1 Introduction

Complicated human behaviour requires the ability to make predictions about *when* events will occur in the environment. There is a long history of treating perception as a the outcome of predictive processes (Gregory, 1968; Helmholtz, 1867), and recently this idea has developed into the increasing influential predictive processing approach (Clark, 2012; Karl Friston, 2005; Seth, 2014). Much previous work has attempted to find the neural correlates of prediction and/or prediction error. One classic example is based on neurophysiological event-related potential (ERP) component called the mismatch negativity (MMN)(Friston, 2005). MMN is a neural response that can be found following experience of infrequent or *unexpected* events in a sequence.

The classical version of MMN was reported using rhythmic sequences of auditory events wherein the majority of presentations, standard trials (e.g. 80%), are of one pitch (e.g. low pitch) and infrequently (20%) a deviant pitch (e.g. high pitch) is presented (references). Under these conditions, a difference in ERP is found such that the waveform is deflected, deviating from the ERP recorded in standard trials (Näätänen, Gaillard, & Mäntysalo, 1978; Näätänen & Michie, 1979; Sams, Paavilainen, Alho, & Näätänen, 1985). Similar findings have also been reported when the deviant is temporal rather than features, such that the rhythm of a sequence of events is disrupted by one event in the sequence occurring earlier or later than expected based on the rhythm – a temporal oddball paradigm (*temporal* MMN; TMMN, Chen, Huang, Luo, Peng, & Liu, 2010; Jacobsen & Schröger, 2003; Näätänen, Paavilainen, & Reinikainen, 1989; Tse & Penney, 2006; see also review Ng & Penney, 2014). TMMN has been reported as a neural response of pre-attentive interval timing processes that track interval regularities and detect deviants (Grimm, Roeber, Trujillo-Barreto, &

Schröger, 2006; Tse & Penney, 2006). However, interpretation of previous TMMN findings is complicated by the use of fixed inter-trial-intervals (ITI) (Hsu et al., 2010; Jacobsen & Schröger, 2003, 2003; Joutsiniemi et al., 1998; Okazaki, Kanoh, Takaura, Tsukada, & Oka, 2006; Roger, Hasbroucq, Rabat, Vidal, & Burle, 2009; Takegata, Tervaniemi, Alku, Ylinen, & Näätänen, 2008; Tse & Penney, 2006; for non-fixed ITIs see also Chen et al., 2010; Grimm et al., 2006; Grimm, Widmann, & Schröger, 2004). Using fixed ITIs, stimulus events across trials inevitably create a rhythmic temporal structure. Recent evidence shows that the neural mechanisms of interval timing (also called *duration-based timing)* can be considered as distinct from rhythmic processing (also called *beat-based timing*) (Grube, Cooper, Chinnery, & Griffiths, 2010; Teki, Grube, & Griffiths, 2012; Teki, Grube, Kumar, & Griffiths, 2011). The transition of rhythmic structures has already been shown to cause mismatch responses in both ERPs (Ford & Hillyard, 1981; Vuust et al., 2011) and event-related fields (ERF) recorded by magnetoencephalography (Vuust, Ostergaard, Pallesen, Bailey, & Roepstorff, 2009). Therefore, in studies that use a temporal oddball paradigm to assess neural correlates of temporal prediction, it could be the rhythmic structures caused through fixed ITIs that drive the MMN response, not neural mechanisms dedicated to predictions of interval timing itself. Considering the evidence for distinct processing of rhythm and duration processing it is reasonable to think that a temporal predictive mechanism may exist independently of rhythmic structure and, consequently, that violations of predictions about temporal interval (non-rhythmic) could also elicit neural processes that may generate mismatch responses. Therefore, TMMN might not be driven by predictions about time itself, but prediction of rhythmic structure alone. To investigate this possibility we compared mismatch responses in isochronous and anisochronous stimulus sequences, in which the ITI was fixed or randomly sampled from a distribution that rules out the contribution of global rhythmic structure. We also examined whether the neural mechanisms underlying temporal prediction are modality-general or specific. Some studies have

suggested that temporal processing might result from modality specific operations (Wearden, Todd, & Jones, 2006). However, accurate temporal predictions may benefit from the integration of temporal information from different sources (Shi, Church, & Meck, 2013). In such a case we would expect to find evidence for similar patterns of neural activation (measured using EEG) elicited by violations of temporal prediction across auditory and visual domains, despite the differences in sensory evoked potentials between the two modalities.

To examine these hypotheses, we used a combination of classic (ERP) and cutting edge machine learning analysis techniques - multivariate pattern analysis (MVPA) and temporal generalisation analysis (TGA; J.-R. King & Dehaene, 2014). We used MVPA to decode neural response patterns to violations of temporal prediction embedded in EEG signals. We then used TGA to investigate whether decodable neural patterns can be used to decode neural activations at other time points both within a given condition, and between conditions that differ in temporal properties and/or sensory modality of presentation. If the topographical patterns between standard and deviant trials are distinguishable across conditions, a binary classifier should be able to classify topographic patterns recorded from different conditions and reach above chance classification performance. The combination of MVPA and TGA allowed us to examine the commonalities and differences of the neural responses related to temporal predictive mechanisms both between isochronous and anisochronous sequences, and between modalities as they evolve over time following unpredicted sensory events.

4.2 Methods

4.2.1 Participants

16 healthy students with normal or corrected-to-normal vision were recruited from the

University of Sussex (8 males, age range 18–31 years). Written informed consent was acquired from all participants prior to the study, which was approved by the University of Sussex ethics committee. Participants received £15 as compensation for their time.

4.2.2 Procedure and Design

Participants were seated in a dimly lit electromagnetically shielded room and asked to maintain fixation ~ 50 cm away from a gamma corrected LaCie Electron blue IV 22" CRT Monitor. Stimuli were generated and presented using the Psychophysics toolbox (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007). Auditory stimuli were played through stereo speakers with an intensity level of ~ 65 dB SPL.

To investigate the processes underlying prediction of temporal intervals, we employed a typical temporal oddball paradigm. In each trial, two transient stimuli were presented, the first stimulus (S1) defining the beginning and the second stimulus (S2) the end of the specified interval (i.e., inter-stimulus-interval, ISI). The interval could be either 150ms or 400ms (Figure 4-1). There were two blocks for each experimental condition. In one block an interval of 150ms in duration was presented in 200 trials (standard) with an interval of 400ms presented in 50 trials (deviant). In the other block, the standard and deviant interval were switched. The position of the deviants within the sequence was pseudo-randomised but constrained so that the deviant was presented at least once in every five trials.

Three experimental conditions were examined in this study, auditory isochronous, auditory anisochronous, and visual anisochronous. In the auditory conditions, the stimuli consisted of 10ms pulses of 1500 Hz pure tones. In the visual condition, the stimuli were 10ms flashes of luminance defined Gaussian blobs against a grey background (Michelson contrast of 1). The inter-trial-interval (ITI, the time between the first event in one pair and the next) in the isochronous

condition was fixed at 1750ms. In the anisochronous conditions, the ITIs were drawn from two uniform random distributions between 1000-1500ms or 2000-2500ms (minimum of 1000ms and max of 2500 ms) so that the average presentation rate was the same as the ITIs in the isochronous condition, but no set of trials ever occurred at that average repetition rate (Figure 4-1).

There was no explicit task performed in this study. Participants were instructed simply to attend to the sequence of events. The order of completion of the six blocks was pseudo-randomised across participants. Each block of 250 trials took approximately 10 minutes to complete and the entire experiment took approximately 1 hour.

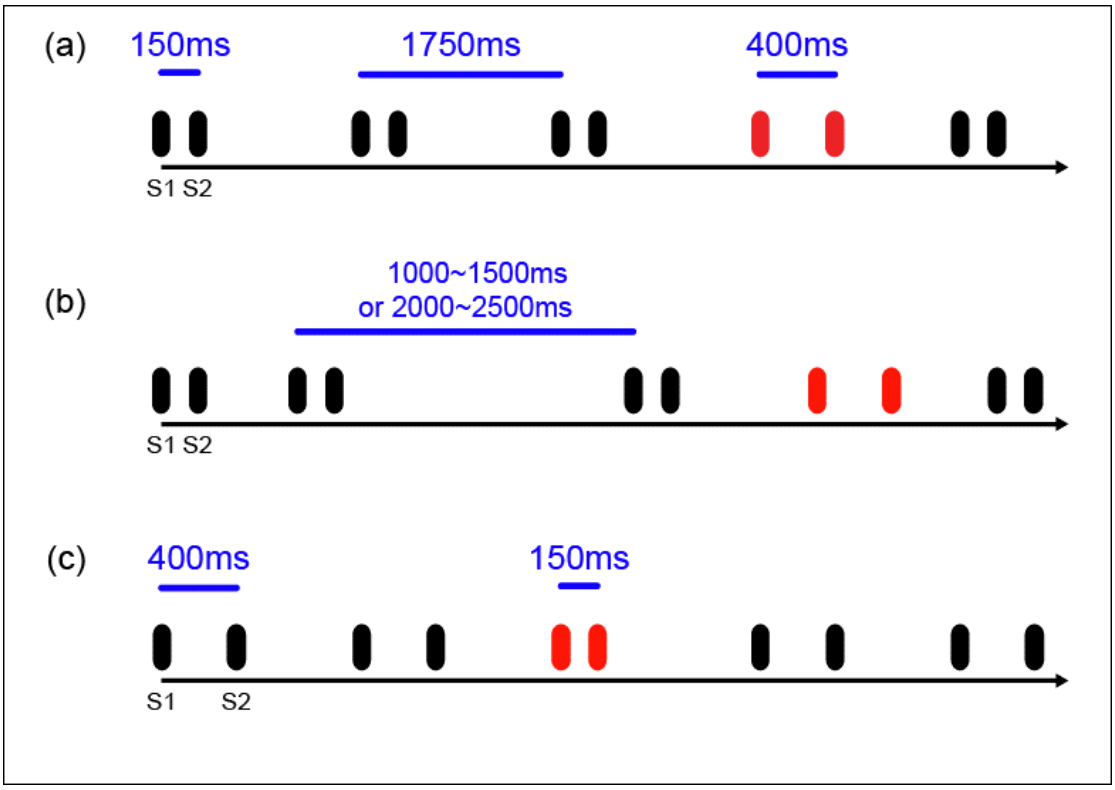

Figure 4-1 Examples of stimulus sequences. (a) An isochronous sequence consisted of short standard intervals (150ms) and long deviant intervals (400ms, depicted by red dots). S1 and S2 denote the first and second transient consecutive stimuli defining a temporal interval. Inter-trial-intervals (ITI) were fixed to 1750ms in isochronous sequences. (b) An anisochronous sequence with varied ITIs. (c) An isochronous sequence consisting of long standard intervals (400ms) and short deviant intervals (150ms).

4.2.3 EEG Acquisition

All EEG data were recorded using a 64 channel ANT Neuro amplifier at a sampling rate of 2048 Hz. A 64 channel Waveguard EEG cap (ANT Neuro, Enschede) employing standard Ag/AgCl electrodes placed according to the 10-20 system was used. Horizontal and vertical eye movements were recorded by two independent pairs of electrodes. The impedances of recording electrodes were maintained below 10kΩ. No analog filter was applied during on-line recording.

4.2.4 EEG Preprocessing

Pre-processing was performed using the EEGLAB toolbox (Delorme & Makeig, 2004) under Matlab (Mathworks, Inc. Natick, MA, USA), and custom Matlab scripts. Continuous data were first down-sampled to 512 Hz. The signals were then band-pass filtered at 0.5-30 Hz and epoched between -140ms and 1400ms relative to onset of the first stimulus in each stimulus pair. Each epoch was baseline corrected automatically identified and removed using the automated Independent Component Analysis (ICA) rejection algorithms ADJUST (Mognon, Jovicich, Bruzzone, & Buiatti, 2011) and MARA (Irene Winkler, Haufe, & Tangermann, 2011). Epochs containing signal values exceeding a threshold of $\pm 75 \mu V$ and deviations greater than by subtracting the signal average of the prestimulus interval (-140-0ms). Artefacts were 6 standard deviations of the mean probability distribution on any single channel were automatically rejected. For the multivariate pattern analyses (MVPA, mentioned below), epochs were further down-sampled to 128 Hz.

4.2.5 ERP Analysis

For each ERP analysis, epochs were averaged separately within each comparison condition for each participant. Statistical analyses of the ERP data were performed at the group level (16

participants, see Statistical Analyses and significance testing)

4.2.6 Multivariate Pattern Analysis (MVPA)

4.2.6.1 Decoding EEG activation patterns between standard and deviant evoked responses

To test whether standard and deviant temporal intervals elicited distinct topographical patterns, we applied MVPA to EEG topography. We reasoned that if the topographical patterns between standard and deviant trials are distinguishable across conditions, a binary classifier should be able to classify topographic patterns recorded from different conditions and reach above chance classification performance. We employed a Linear Support Vector Machine (SVM) classifier (Vapnik, 2013) for every classification analysis. In the current study, an unbalanced trial number between standard and deviant trials (close to 4:1, different trials were excluded during pre-processing) was expected because of the nature of the oddball paradigm. To avoid biased classification due to training on unbalanced data sets, every MVPA analysis was performed on subsampled data sets (Maimon & Rokach, 2005). A subsampled data set consisted of all deviant trials and randomly subsampled standard trials with equal trials number of deviant trials.

For each time point and each participant, we performed the MVPA on the data from 64 channels as features in two steps. First, to find the optimized regularization parameter (C: Cost), a subsampled data set was selected and normalised to z-scores (mean subtracted and divided by standard deviation within features, i.e., electrodes) to examine the optimized C. We searched the parameter space from 2^-3 to 2^3 (exponentially spaced) and evaluated the classification performance (classification accuracy) by stratified ten-fold cross-validation (CV). For each fold of the ten-fold CV the subsampled data set was split into training and testing trials with a 9:1 ratio. An SVM classifier was fit on the training trials (training set) and tested on the testing trials (testing set).

The classification accuracies across ten folds were than averaged as the CV accuracy. The C that maximised the CV accuracy was selected as the optimal C value.

In the second stage, a new subsampled data set was used. The new subsampled data set contained no overlapping standard trials from the first step. We performed the same CV procedure as the first step except that the C was fixed at the optimal value obtained from the first step. The CV accuracy was then computed as the classification accuracy.

We repeated this procedure 50 times with different subsampled data sets in each repetition. The classification accuracies across repetitions were then averaged to obtain a stable and unbiased classification performance across entire data set. All classification analysis were conducted by LIBSVM library (Chang & Lin, 2001) included in e1071 and Caret packages in R software (Team, 2014).

4.2.6.2 Temporal generalization analysis

To examine the possibility that temporal intervals presented in isochronous and anisochronous sequences, and auditory or visual modalities, share some common underlying neural processes, the temporal generalisation analysis (TGA) was used (adapted from King & Dehaene, 2014). The standard decoding analysis mentioned above is adequate to assess whether EEG signals is informative to successfully distinguish standard and deviant evoked activation pattern in a single comparison analysis. However, because information processing may evolve differently between conditions, an analysis approach that considers potential latency difference between condition is needed. For this purpose, TGA was used to evaluate common processing between condition. If a classifier trained on activation patterns at a specific time point can also classify activation patterns at another time point with an above chance accuracy, there should be similar neural processes eliciting the similar topography at the two time points. Therefore, TGA provides an estimation of

common processing and can be used to examine the latency difference of common processes between conditions (J.-R. King & Dehaene, 2014, p. 2).

As for the above described MVPA, the TGA procedure was repeated for 50 runs for each training time point and participant. In each run, a subsampled data set containing an equal number of standard and deviant trials (the training set) for a given training time point (training time, i.e., $t_{training}$) in one condition was trained. This classifier was then used to test subsampled, balanced data sets in which each data set (testing sets) was the data from a time point (testing time, i.e., $t_{testing}$) in another condition. Therefore, a trained linear SVM model from one condition was used to predict the trial types (standard or deviant) at time points in the data from another condition. For example, one can train a classifier using EEG data at 300ms (training set) in auditory anisochronous sequence to classify standard and deviant trials. If the trained classifier is able to successfully predict the trial types using data at 350ms (testing set) in auditory anisochronous sequence, the result suggests that similar decodable information is embedded in the training and testing sets and can be captured by the classifier even from different time points and conditions. This procedure yielded 198 generalisation accuracies (1.54s epoch with 128Hz sampling rate). The same procedure was applied to each training time point (198 training time point in total). A generalization matrix with training time x testing time (198 x 198) was obtained to examine the possible shared neural processes between conditions. Finally, the 50 generalisation matrixes from the 50 runs were then averaged.

In a generalization matrix, an above chance generalisation accuracy on the diagonal (training time = testing time) suggests shared processes occurring at the same latencies between conditions. If a matrix shows an off-diagonal pattern, the shared processes occurs at different timings in the two conditions (J.-R. King & Dehaene, 2014; J.-R. King, Gramfort, Schurger, Naccache, & Dehaene, 2014).

4.2.7 Statistical Analyses and significance testing

Cluster-based permutation analysis with Monte Carlo randomization (Maris & Oostenveld, 2007) was performed for both ERP and MVPA analyses using custom scripts adapted from the FieldTrip (Nichols & Holmes, 2002; Oostenveld, Fries, Maris, & Schoffelen, 2010) and Mass Univariate Toolbox's (Groppe, Urbach, & Kutas, 2011). This method was conducted to reduce the family-wise type I error rate due to multiple comparisons. The algorithm considers an effect is a true positive when a group of adjacent analysis points (neighbours) reach statistical significance together. Two steps were taken in this procedure. First, for every testing point, a statistic across participants was computed (e.g. t-value). A critical value was then used for thresholding testing (e.g. $p < .05$, one-tail). Second, adjacent data points that exceeded the critical value together were defined as a cluster. For each cluster, the cluster "mass" was computed by summing all statistics from each data point within a given cluster. To construct a null distribution from permutations, the same procedure was repeated 5000 times with shuffled condition labels for each data point and participant. The maximum cluster mass from each permutation was then used to construct the null distribution. Finally, the cluster-level p-value was computed by identifying the rank of cluster mass from real data in the null distribution (Monte Carlo p value).

For ERP analyses, the neighbours were defined as adjacent channels and sample time points. The adjacent channels were computed by Delaunay triangulation in 2D projection of the sensor position. In the first step, critical values were computed using a two-tailed dependent t statistic with a $p < .05$ criterion.

For, MVPA analyses for 1D data (time only), the neighbours were defined as adjacent sample time points. For the TGA, the neighbours were defined as adjacent training and testing time points in a 2D space. Because we only considered above chance level classification accuracy, critical

values were computed from one-tail one-sample t statistics again using 50% accuracy with a $p < .05$ criterion.

For the robust multilinear regression analysis, a one sample t-test was performed to test the Fisher's z-score for each time point against an average score calculated from the pre-stimulus period.

4.3 Results

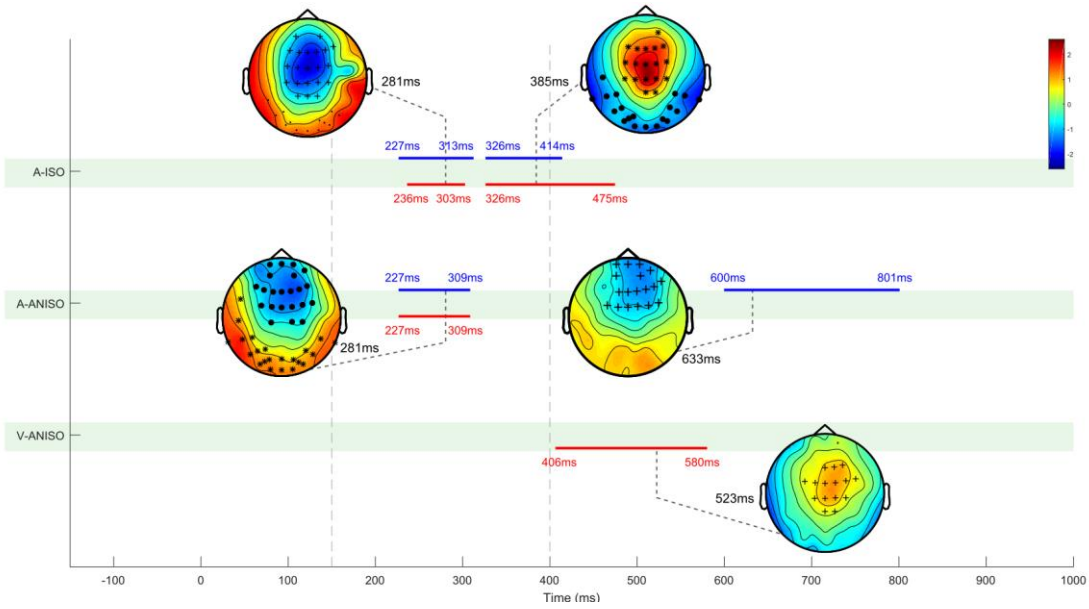

Figure 4-2 Unexpected sensory input elicited mismatch responses. ERP mismatch spatiotemporal clusters by the contrast between standard and deviant intervals with physically identical stimuli (150ms interval). The durations of statistically significant positive and negative clusters are depicted by blue and red lines respectively. The representative topographical distributions of each cluster are selected from representative time within the time range of each clusters. Electrodes marked by symbols indicate spatial ranges of each cluster at the representative time. A-ISO: Auditory isochronous sequence. A-ANISO: auditory anisochronous sequence. V-ANISO: visual anisochronous sequence.

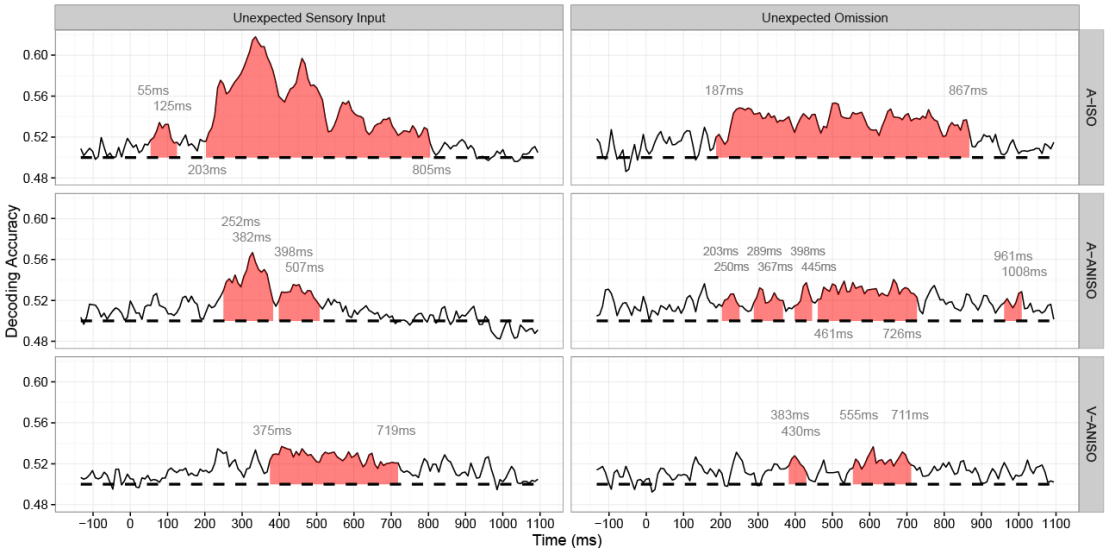

Figure 4-3 Average temporal decoding accuracy across all participants. Classiifiers were trained and tested to classify standard and deviant intervals with data from 64 channels. Highlighted red areas are statisitcally signnificant time windows by cluster-based permutation with 1D (time) data. Unexpected Sensory Input is defined by analyses on trials with short (150ms) stimulus intervals.Unexpected omission is defined by analyses on trials with long (400ms) stimulus intervals. A-ISO: Auditory isochronous sequence. A-ANISO: auditory anisochronous sequence. V-ANISO: visual anisochronous sequence.

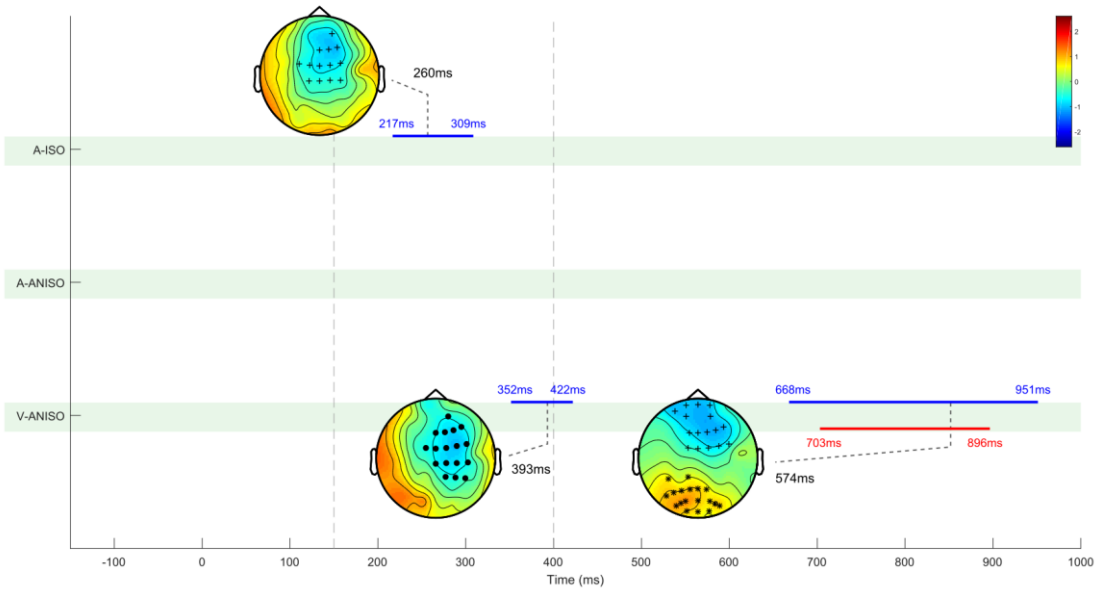

Figure 4-4 Unexpected omission elicited mismatch response. ERP mismatch spatiotemporal clusters by the contrast between standard and deviant intervals with physically identical stimuli

(400ms interval). The durations of statistically significant positive and negative clusters are depicted by blue and red lines respectively. The representative topographical distributions of each cluster are selected from a representative time within the time range of each cluster. Electrodes marked by symbols indicate spatial ranges of each clusters at the representative time. A-ISO: Auditory isochronous sequence. A-ANISO: auditory anisochronous sequence. V-ANISO: visual anisochronous sequence.

4.3.1 Neural correlates of temporal prediction driven by duration processing

4.3.1.1 Evidence from unexpected sensory events

Neural correlates of temporal prediction were first examined by neural responses to unexpected sensory events. In a long standard interval block of trials, participants were presented with the long standard interval (400ms) and a short interval (150ms) deviant. Consequently, the second stimulus (S2) arrived at an unexpected timing (150ms). By comparing the short interval (150ms) with physically identical short intervals in the standard trials (150ms) from the counterbalanced block, we can examine neural responses driven by temporal predictive processing with physically identical sensory stimulation.

We first replicated the results from previous temporal oddball studies in which rhythmic information, due to fixed ITIs (isochronous sequences), also contributed to temporal prediction of sensory events. The results from the ERP comparison between standard and deviant trials with the same short intervals on the auditory isochronous sequences showed two pairs (both significant positive and negative parts within similar time windows) of significant clusters (Figure 4-2). The early paired clusters at 227ms-313ms showed a negative ERP deviation in response to the unexpected auditory input, with a central to frontal distribution ($p = .025$ and .014 for the positive and the negative cluster, respectively). The late paired clusters showed a similar but polarity-reversed distribution at 326ms-475ms ($p = .003$ and .019 for the positive and the negative

cluster, respectively.) These results are consistent with previous findings indicting EEG correlates of prediction for rhythmic temporal sequences (Y. Chen et al., 2010; Jacobsen & Schröger, 2003; Joutsiniemi et al., 1998). Note that, the latency from the second stimulus onset was 77ms (227ms – 150ms) and 176ms (326ms – 150ms) for the early and the late effects respectively. The notably early 77ms latency suggests the temporal prediction influenced neural responses of early auditory processing (Katz, 2014).

We next examined whether EEG patterns across channels contained information about trial types with MVPA. This approach shows better sensitivity than the above univariate ERP analyses. Linear SVM could classify standard and deviant EEG patterns in two time windows with above chance performance. The later time window showed a longer time range from 203ms to 805ms ($p < .002$) than the ERP analysis suggesting that information about trial type (standard or deviant) was embedded across multiple electrode channels and was not captured by traditional univariate analyses.

Surprisingly, we found a remarkably early significant window ranging from 55ms to 125ms ($p = .03$). Note that this time range is earlier than the disclosure of trial types (i.e., 150ms). Therefore, the early EEG pattern difference cannot be attributed to any violation of temporal prediction, but was more likely due to preparation processing for incoming S2 caused by the rhythmic structure of isochronous sequences.

Having broadly replicated previous results reported for rhythmic sequences, we examined whether our anisochronous stimulus sequences that, by design, are free from global rhythmic components could also provide evidence for neural correlates of temporal predictive processing.

For the auditory anisochronous sequences, the ERP analysis (Figure 4-2) revealed a pair of significant early clusters ($ps < .002$ for both positive and negative clusters) and a late negative cluster (600ms to 801ms, $p < .002$). These ERP differences between standard and deviant trials

demonstrate that neural correlates of temporal predictions can be found for duration itself, in isolation from rhythmic temporal information.

Similar to the results from the auditory isochronous sequences, a significant early cluster pair showed that unexpected auditory input elicited a central-to-frontal negative response at 227ms and 309ms. The early onset of these clusters suggests that prediction about temporal interval influenced very early auditory processing occurring only 77ms from S2 onset time (227ms – 150ms = 77ms). In combination with the results from presentation of isochronous sequences, these data indicate that both duration and rhythmic predictions modulated early auditory processing.

However, contrary to the auditory isochronous sequences, ERP analysis of anisochronous sequences did not show evidence for the existence of a second cluster (from around 300ms - 400ms), indicating a unique contribution from rhythmic information to temporal predictive processing. Conversely, we found that a late central-to-frontal negative cluster only in auditory anisochronous sequences, suggesting a unique contribution of prediction about temporal interval distinct from rhythm.

Next, linear SVMs were used to decode EEG patterns. The results show two time windows with above chance decoding accuracies. The early time window is consistent with the early ERP cluster and further extends the time range to 383ms (250ms to 383ms, $p < .002$), suggesting MVPA is more sensitive to the neural response difference between standard and deviant trials with auditory anisochronous sequences. The second time window occurred at 398ms to 508ms ($p < .011$). This result suggests that some neural responses elicited by unexpected auditory input were distributed in a global pattern across multiple electrodes that could not be fully captured by traditional univariate ERP analyses.

With the same ERP and MVPA approaches as applied to the auditory sequences, we also examined the neural correlates of temporal prediction for rhythm-free visual sequences. The ERP

cluster-based analysis revealed unexpected visual input elicited a significant mismatch cluster between 406ms to 580ms ($p < .002$) around central to frontal areas. With MVPA, again, we showed a much longer significant time window with above chance decoding accuracies at 375ms to 719ms ($p < .002$), suggesting that visual prediction about temporal interval was partially encoded by a global pattern of neural response that cannot be fully captured by traditional univariate ERP analyses. Note that, as compared with the auditory sequences, both ERP and MVPA approaches show that the visual mismatch response was much later than auditory one, suggesting a fundamental difference in information processing speed between auditory and visual systems, consistent with what is known about the difference systems (A. J. King & Palmer, 1985; Regan, 1989).

4.3.1.2 Evidence from unexpected omission

Violation of temporal prediction can also occur for the timing at which a sensory input is expected but omitted, rather than being unexpectedly presented. To examine the neural processes elicited by unexpected omissions, we compared neural responses from the long standard trials (400ms) with those in the long deviant trials (400ms) in which a short (150ms) interval was expected and thus S2 was expected to occur 150ms after the S1 onset. In such circumstances, anticipation of S2 would be entirely driven by top-down expectation. We would expect that a certain noise level in the temporal expectation process would lead the anticipated timing of S2 arrival to distribute around 150ms across trials. Therefore, we predict that (i) because the time course of neural response for the unexpected omission was not aligned across trials, a smaller effect was expected in the ERP analysis and the MVPA. (ii) the average neural response driven by unexpected omission across trials should be earlier than by unexpected sensory input. This is because anticipated time for S2 in some trials is earlier than the time of S2 in standard trials

(150ms). However, the neural response for unexpected sensory inputs was always aligned well and triggered after a real S2 at 150ms.

We first examined temporal prediction from isochronous auditory sequences. The ERP analysis showed that unexpected omission elicited only a negative ERP at centre-to-frontal areas from 217ms to 309ms ($p = .023$, see Figure 4-4). The MVPA further showed that the EEG pattern across channels from standard and deviant trials can be decoded from 188ms to 867ms ($p < .002$). As for the findings driven by unexpected sensory input, this result shows that the human neural system can learn temporal regularities and form temporal prediction from rhythmic sequences. Note that both the ERP analysis and MVPA show a very early onset time of the effect which is earlier than the effect driven by unexpected sensory input (227ms and 203ms for the ERP analysis and MVPA, respectively)[Error! Bookmark not defined.].

We next tested whether sequences without rhythmic information can also demonstrate temporal prediction. Even though the ERP based analyses did not show any significant cluster (Figure 4-4), MVPA (Figure 4-3) still showed several significant time windows starting from 203ms to 727ms ($p = .03$, .01, .04, and $< .001$, ordered chronologically) and a very late time window at 961ms to 1008ms ($p = .04$). Again, this result suggests information regarding predictions of temporal interval was embedded in the global neural activation pattern that the univariate ERP analysis was unable to detect. The onset time of the first significant window (203) was again early compared to the effect driven by unexpected sensory input in auditory anisochronous sequence (227ms and 250ms for the ERP analysis and MVPA).

Unexpected visual omission also elicited mismatch responses for visual anisochronous sequences. Here, we found an early ERP cluster (352ms to 422ms, $p = .02$) and a late pair of ERP clusters (from 668ms to 951ms, $ps < .001$). MVPA showed two significant time windows. The first window (383ms to 430ms, $p = .04$) is consistent with the time range of the early ERP cluster. The

second cluster starting from 555ms to 711ms ($p < .001$) partially overlaps with the late ERP cluster, showing that the ERP cluster-based analysis and MVPA both captured unique neural response differences between standard and deviant trials. Compared to neural responses elicited by unexpected auditory omission, the onset of earliest visual response to unexpected visual omissions, again, showed a longer delay from S2 onset time in MVPA (383ms for visual omission vs 203ms for auditory omission). Importantly, this result suggests that the temporal prediction process takes sensory information processing speed into account. Even though the unexpected auditory and visual omissions were physically identical (nothing occurred), a delay was still present between the onsets of auditory and visual mismatch responses. Similar to the finding in the auditory domain, the unexpected visual omission elicited mismatch responses started earlier than the one by unexpected visual inputs in ERP analysis (352ms vs 406ms, but similar in MVPA 383ms vs 375ms).

4.3.1.3 Interim summary

We replicated TMMN responses elicited using isochronous sequences as reported in previous research (Y. Chen et al., 2010; Joutsiniemi et al., 1998; Näätänen et al., 1989, 2007). Examining the neural responses in anisochronous sequences by ERP cluster-based analysis and MVPA, we found neural responses elicited by violations of prediction for temporal interval for both auditory and visual sensory inputs – an explicitly *duration* MMN. Generally, MVPA showed higher sensitivity for EEG differences between standard and deviant trials and could capture information embedded in global patterns of neural response that were not detected by the univariate ERP approach.

4.3.2 Finding cross-modal temporal prediction between audition and vision

Next, we examined whether the neural responses related to predictions about temporal intervals elicited by auditory and visual sequences contained shared information that may be driven

by a supramodal temporal predictive processing system. More specifically, if both auditory and visual deviant intervals were detected by the same supramodal temporal prediction system we expected that we should observe similar neural response patterns embedded within the EEG signal. In such a case, a neural pattern classifier trained to decode EEG signals from auditory standard and deviant trials should be able to decode EEG signals from visual standard and deviant trials. However, as shown above, the information processing speeds of auditory and visual inputs are different (auditory was always faster than visual). To accommodate the perceptual latency between modalities (A. J. King & Palmer, 1985; Regan, 1989), we adopted temporal generalisation analysis (TGA; J.-R. King & Dehaene, 2014).

If a classifier trained on EEG signals at a specific time point (training time) is able to classify EEG patterns at another time point (testing time) with an above chance accuracy, EEG signals between the two time points is likely to contain shared information about trial types (J.-R. King & Dehaene, 2014).

4.3.2.1 Supra-modal predictions of unexpected sensory input

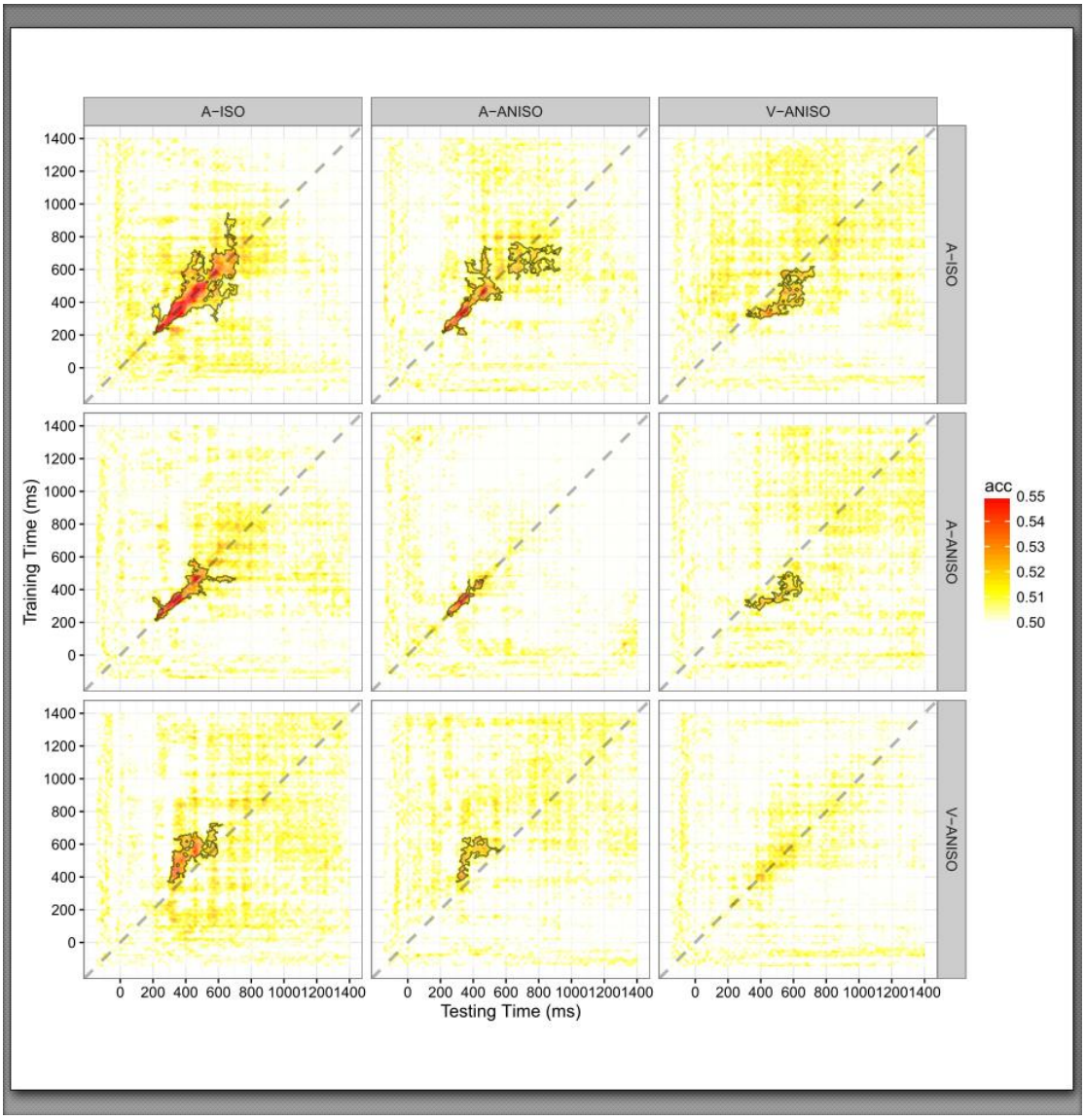

Figure 4-5 Temporal generalisation analysis for unexpected sensory input between conditions. Classifiers were trained on the data at each training time point and tested on data at each testing time point. Three diagonal panels depict cross-validation decoding accuracy for within condition temporal generalisation analyses. The non-diagonal panels illustrate between condition temporal generalisation analyses. Outlines indicate statistically significant clusters. In each panel, along diagonal clusters show EEG signals at close training and testing time points share similar information. Below diagonal clusters indicate that information at training time point occurs in late testing time. Above diagonal clusters indicate that information at training time point occurs in early testing time.

We first examined whether shared information was embedded in neural patterns elicited by the temporally unexpected visual input and unexpected auditory input. At every sample point (training time) in the epoch range (-140ms to 1400ms), linear SVM classifiers were trained to classify neural patterns elicited by *auditory* standard and deviant anisochronous sequences (see methods for details). These classifiers were then used to classify EEG signals at every time point (testing time) in *visual* anisochronous standard and deviant trials. After repeating the training and testing procedure for all combinations of training and testing time, the method created a generalisation matrix. The result shows a below-diagonal cluster with above chance classification accuracies ($p = .021$, Figure 4-5). Importantly, the below-diagonal pattern illustrates that the information used to decode auditory standard and deviant trials can also be found in the later neural response patterns elicited by visual sequences. This result supports the existence of supra-modal temporal interval predictions, triggered by auditory information earlier than by visual information.

Similar results were also found when we reversed the training and testing data sets. That is, we trained the classifiers on EEG signals from visual anisochronous sequences and tested on data from auditory anisochronous sequences. The generalisation matrix showed an above-diagonal cluster ($p = .025$, Figure 4-5), indicating that the information used to decode neural responses in visual standard and deviant trials could successfully classify the neural patterns at earlier time windows.

We quantified the delay between visual and auditory information processing by computing the average delays within the significant clusters. The TGA on anisochronous sequences from auditory to visual and from visual to auditory show an average 131ms delay ($t(399) = 42.52$, $p < 10^{-149}$) and an average 142ms ($t(387) = 46.07$, $p < 10^{-158}$) delay, respectively. Therefore, our results support the existence of a supra-modal temporal predictive process responsible for extracting regularities in temporal interval to generate temporal predictions.

4.3.2.2 Supra-modal predictions of unexpected omission

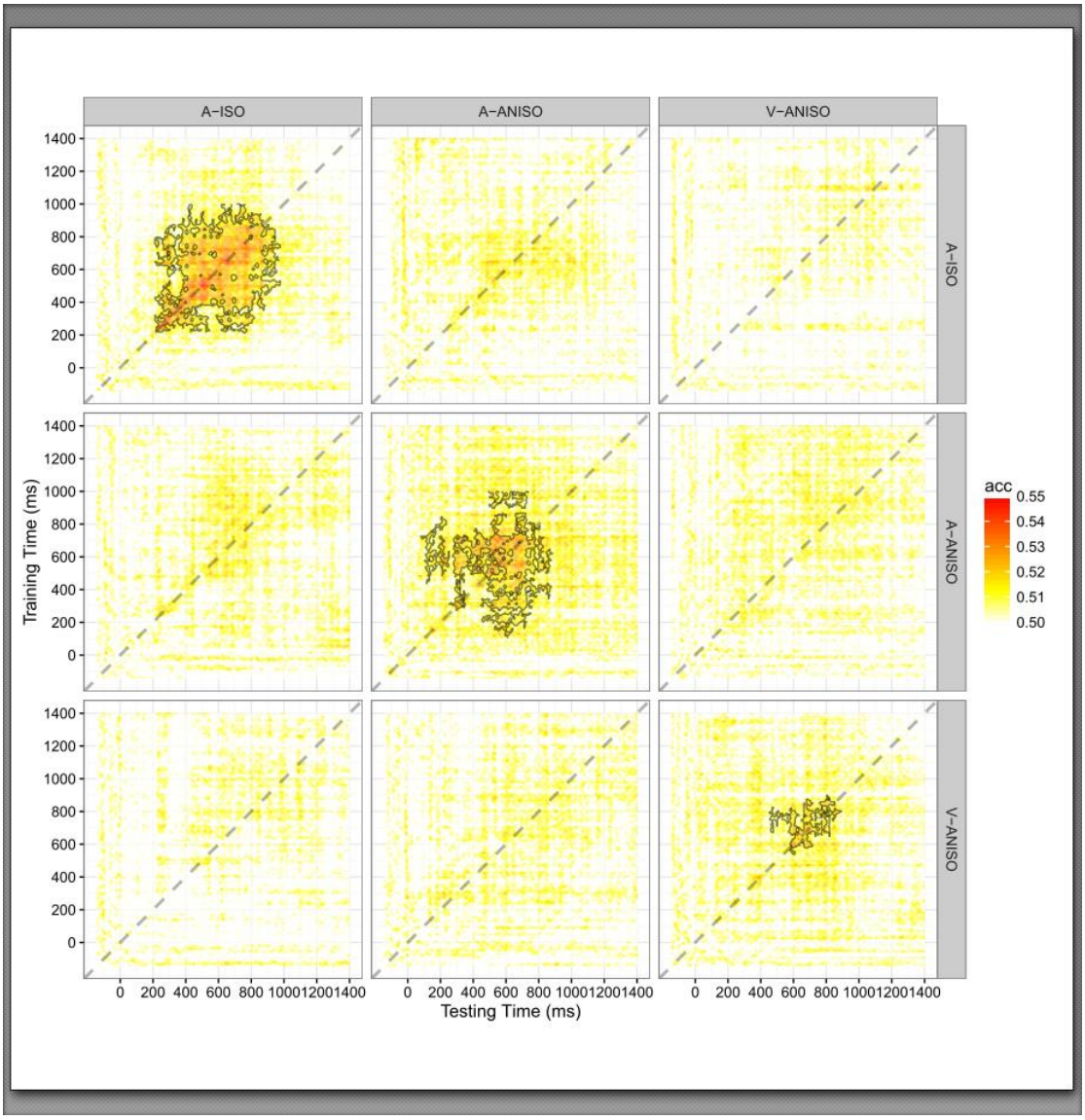

Figure 4-6 Temporal generalisation analysis for unexpected omission

We further examined the shared neural response pattern elicited by auditory and visual unexpected omissions. However, the TGA showed no cluster with above chance classification accuracies (Figure 4-6). This asymmetry between the result from unexpected sensory input and unexpected omission may be due to the impact from internal noise to the expected S2 onset

time[Error! Bookmark not defined.]. The internal noise led the expected S2 onset time generated from internal top-down prediction vary across trials. This resulted in the classifiers being unable to learn a stable neural pattern to classify standard and deviant trials presented in different sensory modality.

4.3.3 Temporal dynamic of temporal prediction processing: different temporal generalisation patterns driven by unexpected presence versus unexpected omission

The detection of unexpected sensory omission relies on top-down temporal predictions of S2 onset time which is inevitably biased by random noise in neural system[Error! Bookmark not defined.]. As a result, the expected S2 onset time may vary across different trials. Therefore, a classifier trained on data at a certain training time point would be able to decode brain patterns (generalisation) in a longer period centred to the training time. For instance, a classifier trained on the data at 250ms would be able to decode EEG pattern at 230ms or 270ms because expected S2 onset time in some trials, due to the internal noise, would be shifted to earlier or later time points. Compared to unexpected sensory omission, unexpected sensory inputs always trigger neural processing precisely at the S2 onset time (150ms). This leads to the neural responses elicited by unexpected sensory inputs being time-locked to S2 onset time and temporally more consistent across trials. Therefore, a decoder would achieve higher decoding accuracies but a shorter range of decoding generalisation time.

We performed the within condition temporal generalisation analysis on the EEG responses from unexpected sensory events and unexpected omission. According to previous experimental and simulation studies (J.-R. King & Dehaene, 2014; J.-R. King et al., 2014), a narrow diagonal pattern in a generalisation matrix represents a series of consecutive but briefly sustained processes. On the contrary, a broad distributed pattern suggests long lasting processes across time. We therefore can examine within condition generalisation matrixes. The results showed that the temporal

generalisation for unexpected sensory inputs in two conditions yielded a narrow diagonal pattern (Figure 4-5Figure 4-4). In contrast, the temporal generalisation for unexpected omission revealed a broader pattern (Figure 4-6). This result indicates that, as we expected, neural processes elicited by unexpected omissions were distributed in a wider time range across trials.

4.3.4 Contribution to temporal prediction from rhythmic information

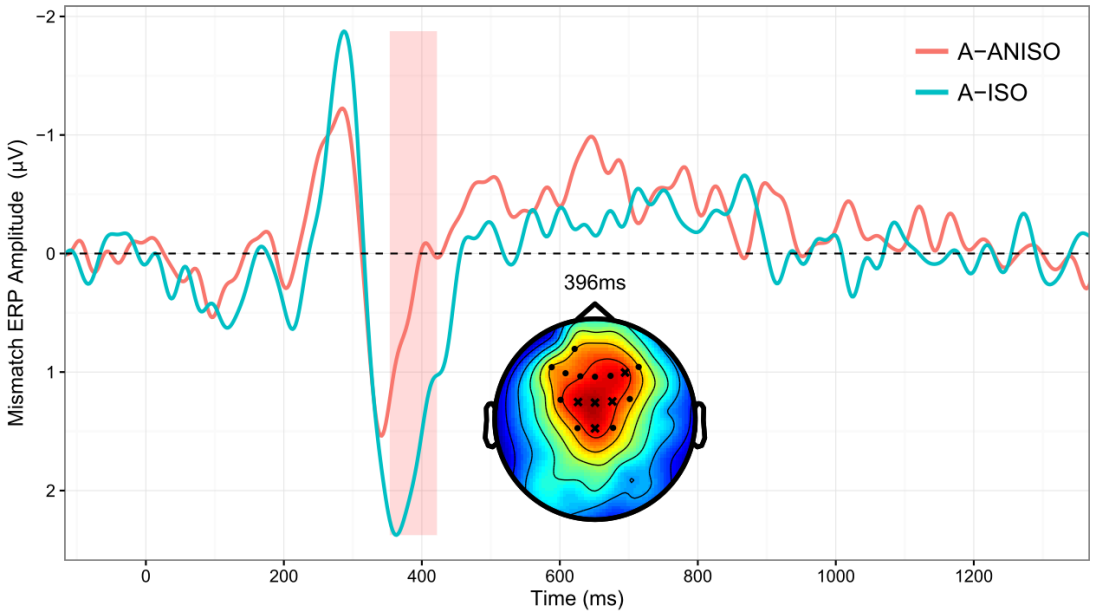

Figure 4-7 Comparison between mismatch responses in auditory isochronous sequence and auditory anisochronous sequence. The red highlighted area indicates the time range of the cluster. ERP waveforms were averaged from electrodes marked by x.

When presenting an isochronous sequence, temporal predictions can be derived from both rhythmic and interval regularities, and therefore, should be more precise than temporal predictions derived from only interval information. Previous research has demonstrated that unexpected auditory events elicit larger neural response amplitudes when predictability of events is higher (Bendixen, Schröger, & Winkler, 2009; Wacongne et al., 2011). We reasoned that violations of

more accurate predictions should lead to a larger prediction error reflected by an increase in response amplitude, but no difference in the temporal dynamics of neural response. We tested this hypothesis by comparing ERP mismatch responses elicited by unexpected auditory inputs between isochronous and anisochronous sequences.

The results showed a significant central-to-frontal positive cluster within a time window around 400ms (354 to 422ms, $p < .01$, Figure 4-7) indicating that unexpected sensory inputs in the isochronous sequence elicited a larger positive mismatch response than in the anisochronous sequence. This result confirms the data presented in the previous ERP analyses indicating that unexpected auditory inputs in isochronous sequences elicit a pair of ERP clusters from 326ms to 475ms that are absent in anisochronous sequences, suggesting a unique contribution from rhythmic information to temporal predictive processing.

More importantly, the temporal dynamic of ERP mismatch response was comparable between isochronous and anisochronous auditory sequences in this time window (Figure 4-7), suggesting that the extra rhythmic information only changed the amplitude of the mismatch response, but not the spatial-temporal dynamic of the neural activation pattern. To test whether similar activation patterns were shared between the two conditions, we performed TGA to examine whether information used to decode EEG patterns in the anisochronous sequence can also be used to decode EEG patterns in isochronous sequences even with the involvement of rhythmic-related processes. The results showed a significant cluster along diagonal when training using data from the auditory isochronous sequence and testing on data from the auditory anisochronous sequence. The decoders were able to classify EEG patterns ($p < .01$) with above chance decoding accuracies. Note that the significant period along diagonal was from 219ms to 523ms, suggesting a similar dynamic of EEG patterns shared between isochronous and anisochronous sequences within this period. The reverse temporal generalisation analysis (training on data from auditory isochronous sequences and testing

on data from auditory anisochronous sequences) also showed a significant cluster with a similar generalisation pattern and time range (diagonal significant period from 219ms to 516ms, $p < .01$), suggesting robust common predictive neural activity shared by auditory isochronous and anisochronous sequences.

4.4 Discussion

Here we examined the neural correlates of temporal prediction using EEG. We first re-examined the temporal oddball design using a combination of univariate (ERP cluster-based permutation analysis) and multivariate (multivariate pattern analysis) approaches. We found that, in the absence of a rhythmic structure in an auditory stimulus sequence, human subjects could still acquire temporal prediction from repeated exposure to a temporal interval. The violation of the temporal prediction yielded a deviation of EEG responses similar to previous findings using sequences that included rhythmic information. These results provide evidence for an explicitly *duration* based process of temporal prediction. We further found similar but delayed neural response patterns induced by visual stimulus sequences as found for auditory sequences, suggesting the involvement of a supramodal temporal prediction mechanism. Finally, we identified unique contributions from duration and rhythmic information to temporal prediction. Together these results suggest the existence of a unified neural mechanism for temporal prediction which integrates temporal information from different sources, including different sensory modalities, to generate an integrated temporal prediction about sensory events.

4.4.1 Temporal prediction from duration and rhythmic processing

Our results are consistent with previous findings suggesting independent neural substrates

underlying duration and rhythmic processing (Teki et al., 2011). We showed that, in the absence of rhythmic information, regularity of duration was sufficient to elicit neural responses to prediction violation. Sequences containing both duration and rhythmic information elicited larger ERP responses than sequences with duration information only, implying the unique contribution from rhythm. This finding agrees with studies suggesting a dissociation between *duration-based* and *beat-based* timing (Grube, Cooper, et al., 2010; Grube, Lee, Griffiths, Barker, & Woodruff, 2010; Teki et al., 2012, 2011).

4.4.2 Duration predictions are processed supramodally

Our results showed that decoders trained on data from one modality (auditory or visual) could successfully decode EEG patterns elicited by deviant intervals from the other modality, suggesting shared neural processes for temporal prediction between the two modalities. It has long been debated whether duration information is processed in a modality specific or a modality-general manner (Buonomano & Karmarkar, 2002; Grondin, 2010; Mauk & Buonomano, 2004). The modality-specific hypothesis suggests that the representation of temporal information is encoded independently and specific to the sensory modalities (Bueti, Bahrami, & Walsh, 2008; Droit-Volet, Meck, & Penney, 2007; Ivry & Schlerf, 2008; Lustig & Meck, 2011; Penney, Gibbon, & Meck, 2000; Takahashi & Watanabe, 2012). The supramodal hypothesis suggests that duration exists as an abstract feature and is encoded in a supramodal/amodal representation (N'Diaye, Ragot, Garnero, & Pouthas, 2004; Rammsayer & Ulrich, 2005; Shih, Kuo, Yeh, Tzeng, & Hsieh, 2009; van Wassenhove, Buonomano, Shimojo, & Shams, 2008; Wassenhove, 2009; Wearden et al., 2006). Using the temporal generalisation analysis, we showed similar, though delayed, neural response patterns elicited by the presence of unexpected sensory information for auditory and visual stimuli. That a similar pattern was found at the delayed time point indicates that latencies of information

processing between modalities should be taken in to account when comparing neural responses from different modalities, something typically not easily addressed in standard ERP analyses of MMN.

4.4.3 Relation of duration MMN to previous MMN

Our ERP cluster-based analyses showed that violations of temporal prediction caused by unexpected auditory inputs in both isochronous and anisochronous sequences elicited an earlier central-to-frontal negative cluster. The topographical distribution and the significant time window of this cluster is consistent with previous reported temporal MMN typically found in auditory temporal oddball paradigms with peak at 100~250 ms after S1 onset or 50~150 ms after S2 onset (Y. Chen et al., 2010; Joutsiniemi et al., 1998; Näätänen et al., 1989, 2007). This result suggests that temporal MMN identified in previous research using temporal oddball paradigms was not merely triggered by temporal prediction from rhythmic structure but also contributed to by temporal prediction from interval timing. A similar cluster was also elicited by unexpected auditory omission in isochronous but not in anisochronous sequences. The weaker temporal MMN-like activation elicited by unexpected omission may due to internal noise[1].

In addition to the temporal MMN-like cluster, violations of temporal prediction in the auditory isochronous sequence elicited a central-to-frontal positive cluster which was absent in the anisochronous sequence. The topographical distribution and the significant time window (176 ms to 325 ms after S2 onset) of the later cluster was comparable with P2 auditory ERP component with a latency of approximately 150–250 ms (Crowley & Colrain, 2004; Katz, 2014). P2 was found to be modulated by deviant auditory stimuli in oddball paradigms and was considered an index of stimulus classification processes (Canales-Johnson et al., 2015; Novak, Ritter, & Vaughan, 1992).

Using auditory anisochronous sequences, unexpected auditory events elicited a late frontal negative cluster. The time window and polarity was close to the reorienting negativity (RON) reported in a previous study using the temporal oddball paradigm (Roeber, Widmann, & Schröger, 2003). In that study, participants were asked to discriminate spatial location of auditory stimuli and ignored infrequent changes of stimulus presentation duration. The result showed a larger RON elicited by infrequent duration (deviant) suggesting a re-orientation from the detected change to the original attentional focus. This RON-like component in the current study was only present in anisochronous auditory sequences, suggesting that the rhythmic structure in the isochronous sequence may serve a function in maintaining participants' attention to the stimulus sequence and preventing distractions from temporally unexpected events.

Unexpected visual inputs elicited a central-to-frontal positive cluster. The topographical distribution and time window (256-430 ms after S2 onset) is close to the p300, more specifically p3a, component found using passive oddball paradigms and considered an index of an automatic orienting to novel and salient stimulus features (Polich, 2007). Interestingly, instead of this positive cluster, we found that unexpected visual omission elicited an earlier central-to-frontal *negative* cluster suggesting distinct neural activities during computations of prediction errors caused by actual visual inputs and absences.

Unexpected visual omissions in the visual anisochronous sequence elicited a late cluster pair. To our knowledge, no previous study has reported similar late ERP components using visual temporal oddball designs. The very late component may reflect higher-order cognitive processes, e.g. attention reorientation, sensory updating, or error correction (Horváth, Roeber, & Schröger, 2009; Schröger & Wolff, 1998; Yeung, Botvinick, & Cohen, 2004). Further research is needed to elucidate the functional rule of this effect in predictive processing.

4.4.4 Auditory and visual information differ in duration prediction access time

Our data showed a longer neural response latency for unexpected visual than auditory events. This suggests that visual information takes longer to access temporal predictive processes than auditory information. When examining the latency difference (the onset of the first significant time window) from unexpected sensory input, ERP analyses and MVPA showed 179ms (406ms – 227ms) and 125ms (375ms – 250ms) delays between visual and auditory evoked responses, respectively. The temporal generalisation analyses on data from auditory and visual sequences showed averages of 131ms (trained on auditory and tested on visual data) and 142ms (trained on visual and tested on auditory data). These results indicate that access to temporal predictive processing for visual information was, at least, 125ms slower than auditory information. This difference was longer than previous studies showing that brain responses occur roughly 30ms to 50ms earlier for the auditory signal (A. J. King & Palmer, 1985; Regan, 1989). Therefore, the larger difference gives a rough estimate of the difference in processing time (~100 ms longer for vision) for generating predications for auditory and visual duration information. However, it is important to note that, in the current design, we did not make any attempt to equate the saliency and intensity level of the visual and auditory stimuli, properties which have been shown to affect sensory information processing speed (Nissen, 1977; Töllner, Zehetleitner, Gramann, & Müller, 2011). Further research is needed to evaluate the impact of stimulus saliency on duration prediction processing speed across sensory domains. Future studies should test this idea more thoroughly by examining whether temporal predictions consider knowledge about processing latencies when standard and deviant durations are presented in different modalities, for example, presenting the standard duration in audition and the deviant in vision, or vice versa.

Interestingly, when examining the difference in neural response latency between unexpected

visual and auditory omissions, the result showed a similar latency difference of 180ms (383ms – 203ms), even though the inputs were physically the same (no input presented). This result is consistent with previous studies on unexpected omission (NITTONO, 2005; Simson, Vaughan Jr., & Walter, 1976), suggesting that neural processing of temporal prediction takes modality-specific latencies of information processing into account.

4.4.5 Neural response to unexpected omission

In the current study, we found asymmetrical neural responses between unexpected sensory inputs and unexpected omissions. This finding replicates previous studies in finding a difference between "increment" (deviant intervals are longer than standard intervals) and "decrement" (deviant intervals are shorter than standard intervals) temporal MMN. (Takegata et al., 2008). A parsimonious explanation is the presence of internal noise within the temporal predictive processing mechanisms. In the current experiment, the unexpected sensory inputs were perfectly time-locked 150ms after S1 onset time. Therefore, neural responses elicited by unexpected sensory inputs would be expected to be triggered at similar timing across trials. However, the onset time of neural responses to unexpected omissions must be determined by only *predictions* for S2 onset time. Consequently, by contrast with unexpected sensory inputs, the onset time of neural responses to unexpected omissions is more susceptible to noise within the predictive processing mechanisms generating predictions of S2. This may result in larger variance in the time course of the neural responses across trials. Our data showed convergent results supporting this possibility. First, we expected that temporal prediction of S2 onset time would occur earlier in some trials and lead to earlier onset time due to the internal noise[1]. Both our data and the data from Takegata et al (2008) showed, in most analyses, that the onset time of the mismatch response to unexpected omission was earlier than for unexpected sensory input (see Table 1 and Table 1 in Takegata et al 2008). Second,

we expected a smaller effect size for unexpected omission than unexpected sensory input due to the variance of temporal predictions of S2 onset time across trials. Both our data and Takegata et al 2008 are consistent with this prediction (see Table 1 in Takegata et al 2008). Third, we directly examined an internal noise account using the within condition temporal generalisation analyses. The results showed that temporal generalisation patterns were notably broader in unexpected omission than in unexpected sensory input. This result indicates that the neural response pattern elicited by unexpected omissions can be found at other time points within a broader temporal window, suggesting less temporal specificity of neural response to unexpected omissions across trials. In sum, our results suggest that the neural processes that rely only on internal prediction to detect prediction violations are more vulnerable to internal noise and cause larger temporal variation in neural dynamic across experimental trials.

Standard vs Deviant

	ERP		MVPA	
	unexpected sensory input	unexpected omission	unexpected sensory input	unexpected omission
A-ISO	227	217	203	188
A-ANISO	227	N.A.	252	203
V-ANISO	406	352	375	382

Table 1 The first significant time points of neural responses elicited by violation of temporal predictions in ERP analyses and MVPA

4.4.6 Null results in this study

The null results I presented in this chapter should be interpreted with caution. A more reasonable way to present these null results is to provide Bayes factors. However, because all the statistical tests in this chapter were performed by the cluster-based permutation analysis with Monte Carlo randomization, I cannot derive an intuitive way to compute Bayes factors. Future research is needed to develop Bayesian statistics for cluster-based permutation analysis.

4.5 Conclusion

Complex human behaviour requires the ability to make predictions about the temporal properties of events occurring in the environment. Our findings demonstrate that humans can acquire and maintain predictions specifically related to duration, in the absence of rhythmic information. We further showed that similar processing components are present when a prediction of duration is violated regardless of presentation modality, indicating that processing of prediction for duration occurs supramodally. Our data suggest that duration as a property of sensory events can be encoded and used to generate predictions, and further, that the associated temporal predictive mechanisms are shared across sensory modality, in line with a general predictive processing framework for perception.

Chapter 5
General Discussion

In this chapter, I will first summarise the results of my three studies (Chapter 2, 3, 4) and interpret them in my three key research questions according to the predictive processing framework. I will move on to discuss the time course of regularity learning and neural signatures of predictive processing in an integrated view across the three current studies. In the final section, I will discuss limitations of this thesis.

5.1 Overview of findings

5.1.1 Cross-modal prediction changes the timing of conscious access during motion-induced blindness

Chapter 2 investigated the relationship between predictive processing and conscious content. I examined whether predictions of visual events by auditory cues facilitated conscious access to visual stimuli. I trained participants to learn associations between auditory cues and colour changes. I then asked whether congruency between auditory cues and target colours would speed access to consciousness. I did this by rendering a visual target subjectively invisible using motion-induced blindness and then gradually changing its colour while presenting congruent or incongruent auditory cues. Results showed that the visual target gained access to consciousness faster in congruent compared to incongruent trials; a control condition excluded the potential confounds of attention and motor responses influencing the results. The expectation effect was gradually established over successive blocks suggesting a role for extensive training in accelerating conscious access. These findings indicate that predictions learned through

cross-modal training can facilitate conscious access to visual stimuli, lending support to predictive processing frameworks of perception (Clark, 2012; Friston, 2005; Rao & Ballard, 1999). Within the context of these predictive processing frameworks perceptual content is determined by probabilistic inference of the most likely external causes of sensory signals. The data from this Chapter supports the view that conscious access occurs when predictive models are verified against sensory inputs leading to the prediction errors being minimized. The results are also consistent with previous findings showing that valid expectation can accelerate conscious access of supressed visual contents using continuous flash suppression (Pinto et al., 2015) and shift a neurophysiological signature of conscious awareness to an earlier time point (Melloni et al., 2011). I suggest that effective predictive cues in this study facilitated the validation of predictive models with sensory signals, leading to more rapid conscious access for this sensory information. Furthermore, the present results extend the predictive processing framework by showing how predictive influences on perception develop across time due to training, and by underlining the flexibility of these influences by demonstrating their efficacy using cross-modal arbitrary associations. Together these results suggest that predictive influences may permeate and shape conscious experiences more deeply and broadly than previously thought.

5.1.2 Visual Perceptual Echo Reflects Learning of Temporal Regularities in Rapid Luminance Sequences

In Chapter 3, I investigated predictive processing for low-level visual features via unconscious visual sequence learning. A sign that the visual system actively processes and predicts sensory signals is revealed by the recently discovered ‘perceptual echo’, in which the cross-correlation between a sequence of randomly fluctuating luminance values and electrophysiological (EEG) signals exhibits a long-lasting periodic (~100ms cycle) reverberation of the input stimulus

(VanRullen & Macdonald, 2012). I hypothesised that the perceptual echo may reflect a periodic process associated with regularity learning using predictive processing. In two experiments and a simulation, I investigated the functional relevance of the perceptual echo response, testing the hypothesis that it reflects a predictive processing mechanism which can encode and learn dynamic visual sequences within the visual cortex.

In Experiment 1, I found that the perceptual echo response was enhanced by repetitions of an identical rapid luminance sequence, indicating that the information about such sequences is encoded by the visual system. Control analyses using shuffled data excluded the possibility that increases in perceptual amplitude could be attributed to general changes in induced alpha-band EEG responses resulting from sequence repetitions.

Experiment 2 first replicated the main finding of Experiment 1, that echo amplitude increases with successive presentations of a random dynamic luminance sequence, and confirmed that the perceptual echo reflected specific sequence information. Following 4 presentations of a specific sequence I compared the echo response elicited by a 5th presentation with that elicited by an 'inverse' sequence, which preserved all non-sequential spectral and temporal properties of the original sequence. Strikingly, the echo amplitude for the inverse sequence returned to a level comparable to presentation of a novel sequence, while a 5th presentation of a non-inverse sequence continued to elicit a strong echo response. I also found that information about an encoded sequence persisted over 9 seconds and was robust to intervening visual input, by showing that the echo response to a re-presented (non-inverse) sequence, following an inverse sequence, recovered to a level indistinguishable from a 6th successive presentation of a given luminance sequence.

To investigate the perceptual echo further I created a predictive processing model, inspired by the Kalman filter (Kalman, 1960), to simulate key features of the perceptual echo. This model provided a computational account of how a predictive coding system may generate the perceptual

echo and also provided insights into the temporal dynamics and morphology of the echo response. The results of this study results extend those of VanRullen & Macdonald (2012) by providing a functional account of the perceptual echo response, demonstrating that it reflects a neural index of the encoding of fast-changing dynamic luminance sequences.

Altogether, this data indicates that human visual areas are capable of rapidly encoding fast-changing dynamic stimuli that appear to be governed by the predictive processing framework, this process being reflected by the perceptual echo. This information persists in the visual system for at least 9 seconds, even in the presence of novel intervening sensory input.

5.1.3 The neural correlates of temporal predictive mechanisms based on interval timing

Chapter 4 investigated the neural correlates of temporal predictive processing in a temporal oddball design using EEG. The interpretation of previous findings using temporal oddball designs to study temporal predictions about interval is complicated due to the use of fixed inter-trial-intervals (ITI), which inevitably confound temporal predictions based on rhythmic information with those based on duration *per se*. I hypothesized that violations of predictions about temporal intervals (non-rhythmic) could also elicit neural responses related to temporal (rhythmic) violations.

With ERP analyses and MVPA, I found that without rhythmic information in the sequence structure violation of prediction about temporal intervals still elicited neural responses related to temporal violations. This finding suggests the existence of specific neural correlates of temporal prediction that are based on interval timing. In addition, I also found a unique ERP component related to rhythmic sequence structures. I further found a central-to-frontal positive ERP component with a larger but similar amplitude when the sequence contained both interval and rhythmic information, compared to sequences with only an interval prediction. This finding

suggests a unified predictive mechanism that integrates both interval and rhythmic information to make more precise temporal predictions about incoming events.

I further investigated whether the neural mechanisms for temporal prediction are modality specific or supra-modal. Using temporal generalisation analysis (J.-R. King & Dehaene, 2014), I found that the neural response patterns elicited by violation of predictions about auditory temporal intervals at earlier time points were similar to those elicited by violation of prediction about visual temporal intervals at later time points. Even though the information processing speeds are very different between these modalities, the results from the temporal generalisation analysis (TGA) that predictive information from auditory and visual temporal intervals can be acquired through a supra-modal predictive mechanism and can be used to generate predictions about both auditory and visual events.

In summary, combining traditional univariate ERP analyses with MVPA and TGA, I successfully identified neural correlates of prediction for temporal intervals (independently of rhythm) and revealed the supra-modal nature of the temporal predictive mechanism. In general, with multivariate approaches, neural responses to the violation of temporal predictions can be detected in longer time windows suggesting that temporal predictions that influence neural processes are more profound than previously thought.

5.2 Thesis research questions:

5.2.1 Predictive processing, conscious access and unconscious inference

Overall, the experimental results of Chapter 2 and Chapter 3 support the notion that neural computations of predictive processing mainly operate unconsciously. This finding fits traditional interpretations of predictive processing, an implementation of the Bayesian brain hypothesis,

which generally posits “perception as unconscious inference” (Gregory, 1980; Helmholtz, 1866; MacKay, 1956; Neisser, 1967; Rock, 1983). According to this view information processing and computation for perception occurs at the unconscious level. Only the final decision of statistical inference reaches conscious awareness and forms the content of conscious experience. In Chapter 2, visual information was supressed during the MIB state. However, I found that cross-modal predictive information can nevertheless facilitate conscious access of visual information supressed under the MIB state, suggesting an unconscious integration between cross-modal predictive information and supressed visual information. These results confirm traditional predictive processing concepts, that information integration and computations associated with perceptual inference operate at the unconscious level, with only the final result of this inference being accessible consciously.

Note that Chapter 2 only examined the impact on conscious access from predictive information learned from a simple cross-modal association. Whether other types of predictive information (e.g., information represented at higher levels in the sensory hierarchy) are also processed and computed at the unconscious level is still unknown. I will discuss this limitation in the following *Limitations and Future Research* Section.

The MIB study was also designed to assess how cross-modal predictions can influence the timing of access to subjectively suppressed (unconscious) visual information. Previous studies that have investigated how valid predictive cues accelerate conscious perception across modalities have focused on the predictions between long-term representations, e.g. semantic-object identity (Y.-C. Chen et al., 2011; Lupyan & Ward, 2013), audio-visual speech (Alsius & Munhall, 2013; T. D. Palmer & Ramsey, 2012b) or motion integration (Conrad et al., 2010). For example, Conrad and colleagues show that, in binocular rivalry, motion sounds can stabilise and lengthen dominance periods for congruent visual motion (Conrad et al., 2010). Similarly, Chen, Yeh, and Spence found

that auditory cues shortened the suppression period of congruent visual objects (Y.-C. Chen et al., 2011). In contrast to these findings, the present study further showed that predictive information learned through short-term, flexible, and learned associations demonstrated a similar impact on conscious perception. Taken together, these findings suggest that cross-modal predictive information exerts substantial influences on unconscious statistical inference, conscious perception, and their interaction.

I theorized that the validity of the predictive information provides different prior information for visual perceptual inference and therefore should affect the inference process. Valid predictive information should accelerate conscious access as the result of unconscious inference, biasing the inference to rapidly arrive at the "correct result" and therefore minimizing the prediction error. Supporting this theory, I found that predictions that were in line with sensory evidence accelerate conscious access. This finding is comparable with previous research showing that perceptual decision criteria shift in different directions based on congruency between perceptual expectations and actual sensory inputs, suggesting decision bias in unconscious inference (Sherman et al., 2015).These results extend the theoretical knowledge about how predictive information affects conscious access by showing how predictive influences on unconscious perceptual inference develop across time as the result of training. .

In Chapter 3, I found that in two experiments, all participants were unaware of repetitions of luminance sequences although the amplitude of the perceptual echo was still able to reveal the effects of sequence learning. From a predictive processing perspective, after the first presentation of a sequence, an interpretation might be that the specific sequence information is encoded as prior knowledge and modulates the perceptual echo in subsequent presentations of the same sequence. The fact that participants failed to explicitly recognise sequence patterns suggests that the sequence knowledge (prior) and the computations underlying the perceptual echo were

unconscious in nature. Furthermore, we observed a peak echo response at approximately 400ms post stimulus onset, suggesting that the visual system was able to detect repetitions of a specific sequence extremely rapidly, within half a second. This result points to the rapid integration between prior knowledge and current sensory signals that operates at an unconscious level. Additionally, we found that the encoded sequence information can persist over many seconds even in the presence of additional intervening sensory input. This finding highlights a number of interesting observations, firstly that sequence specific information (prior knowledge) had been encoded by the visual system unconsciously, an observation supported by the fact that the results of unconscious inference are still available over 9 seconds after the original sequence was presented. Secondly, that the results of unconscious inference between sensory input and prior occurred in this study extremely rapidly, by approximately 400ms. Note that due to the limitation of the current design, this study did not assess the impact from the rapid unconscious inference on conscious perception of the luminance changes. Further research is needed to examine how prior knowledge sequence patterns influences conscious luminance perception.

A central question of this thesis has been to explore the relationship between unconscious statistical inference reflected by predictive processing and conscious access. This thesis has added to our understanding of this question by demonstrating that predictive information is processed unconsciously, and that this information can influence unconscious inference speeding conscious access.

5.2.2 Differing levels of information within the predictive processing framework

As mentioned in the introduction hierarchical predictive coding is capable of encoding a wide range of information types from concrete to abstract. This framework predicts that in humans both low and high level information should operate through the same predictive processing mechanisms.

The studies in this thesis were designed to manipulate the level of information presented and examine predictive mechanisms across these different levels of sensory information. Chapter 3, used low-level sensory information (luminance), Chapter 2 used cross-modal information (auditory-visual) while Chapter 4 utilized supra-modal information (temporal duration).

In Chapter 3, I found increases in echo amplitude with successive repetitions of a sequence demonstrating that participants were implicitly learning information about each sequence, which may reflect perceptual predictions (priors) being updated with the presentation of each sequence. The results of modelling this process (Section 3.3) suggests that the echo response reflects an iterative process that updates priors about the luminance dynamics of a sequence, communicating perceptual predictions at rate defined by the alpha frequency band. Taken together the findings of Chapter 3 support the idea that predictive processing may operate using low-level visual information.

Chapter 2 demonstrated that a visual target gains access to consciousness faster when cross-modal predictions are congruent with the supressed visual inputs. The result is in line with the view that predictive processing in the neural system is organised in a hierarchical manner in which high-level units integrate information across modalities and generate top-down predictions about modality-specific sensory data (Altieri, 2014; Arnal, Wyart, & Giraud, 2011). Previous research on Bayesian brain hypothesis has emphasised the importance of cross-modal perceptual inference (Deneve & Pouget, 2004; Ernst & Banks, 2002; Körding et al., 2007; Shams & Beierholm, 2010). Because an external event typically causes multimodal information which can be perceived by multiple sensory channels, perceptual inference should consider information across modalities to accurately infer the sensory cause. In the predictive coding framework, information from one modality can be used to establish prior of sensory causes and generate precise predictions about sensory inputs from another modality. Therefore, prediction about sensory inputs

can be effectively constrained (Alais & Burr, 2004) by cross-modal predictive information. The current result suggests that cross-modal predictive processing exists in the sensory hierarchy and facilitate perceptual inference by using predictive information from different modalities.

Chapter 4 showed that when an infrequent temporal duration was presented either by visual or auditory modalities the surprise EEG response pattern contained similar and decodable topographical information, suggesting a common predictive mechanism tracking temporal duration regularities across modalities. The result point to a high-level predictive mechanism which operates on abstract supra-modal temporal information that is not modality specific.

Collectively, the results of this thesis answer one of my primary research questions by showing that predictive processing appears to operate across differing levels of sensory information, low-level, cross-modal, and supra-modal.

5.2.3 Regularity learning within the predictive processing framework

How are differing forms of regularity learning used to make predictions about future events within the predictive processing framework? This thesis examined predictive processing via three common types of regulatory leaning utilized in previous studies from psychology and neuroscience. Chapter 2 focused on predictive processing using extensive training to establish predicative relationships between cross modal information. The results showed that cross-modal predictive cues accelerated time to conscious access – but only after extensive training associating specific visual and auditory stimuli. I found a linear relationship between the amount of training and the degree of facilitation, highlighting that the learning of predictive associations can regulate their influence on the timing of conscious access. The results showed how predictive associations develop across time as the result of extensive associative training and how this influences conscious perception. Note that, as discussed in Chapter 2, although this training is extensive it is

still much shorter as compared to the long-term bound representations, e.g. semantic-object identity and audio-visual speech.

The results of Chapter 3 showed that the human visual system is capable of rapidly and automatically encoding and learning temporal regularities embedded within fast-changing dynamic stimulus sequences. This study showed an enhancement of the perceptual echo due to sequence repetition, suggesting the involvement of predictive processing in low-level sequence learning.

In Chapter 4 I identified the neural correlates of predictive mechanisms that tracks temporal duration regularities. The temporal predictive mechanisms were able to acquire the predictive temporal information from duration repetition in a duration oddball paradigm. Based on the temporal information it was able to make temporal predictions about future sensory inputs. Previous studies on temporal duration using duration oddball paradigm have always suffered from a fundamental confound of having two types of temporal information embedded within the paradigm, rhythmic and duration information. This experiment successfully isolated and identified the neural responses relating to temporal duration tracking within this paradigm, independently of rhythmic information. The findings of this study suggest that predictive mechanisms are involved in detecting, extracting, and learning regularities from repetitions of abstract sensory information in the environment, in this case the temporal duration between sensory events.

Together these studies suggest that predictive mechanisms use distinct types of regularity learning to generate valid predictions about future events.

5.3 *The time course of regularity learning and conscious access within predictive processing*

The EEG results of Chapter 3 and Chapter 4 showed that the neural signatures of sequence and

duration regularity learning can be detected within seconds. Chapter 3 showed that the echo enhancement due to sequence repetition linearly increased and rapidly became observable, reaching significance after 4 successive presentations of the same sequence. My data suggests that the regularities encompassing a given sequence was encoded by the 4th repetition (around 24.5 s). This finding is compatible with the learning of random noise sequences in the auditory domain (Agus et al., 2010). In their study Agus et al., demonstrate that the human auditory system can rapidly learn temporal sequence (1s) properties of a random and meaningless input signal (Gaussian random noise) within 10 repetitions (Agus et al., 2010).

Chapter 4 showed that the neural responses of temporal prediction violations were able to be detected using a sequence with 4:1 standard and deviant duration ratio which has often been used in previous MMN studies (Näätänen, Pakarinen, Rinne, & Takegata, 2004). Previous research using the odd-ball paradigm has shown that the MMN elicited by a deviant stimulus is detectable after just a few repetitions of the standard stimuli (Baldeweg, Klugman, Gruzelier, & Hirsch, 2004; Haenschel, 2005). For example, the MMN latency decreases with only two repetitions of deviant stimuli after over 30 repetitions of standard stimuli (Liégeois-Chauvel, Musolino, Badier, Marquis, & Chauvel, 1994). These results suggest that the mechanism of regularity learning underlying MMN can be elicited with just a few repetitions.

Together, the regularity learning of sequence repetition in Chapter 3 and acquisition of temporal information from duration repetition in Chapter 4 suggests a short-term, flexible, and adaptive mechanism that operates in the timescale of seconds. However, Chapter 2 showed that the facilitation of conscious access caused by predictive information was gradually established over a much larger time-scale using cross-modal associative training, suggesting that the impact on conscious access is through a long-term (at least several minutes) regularity and extensive (tens to hundreds samples) learning process.

What are the possible reasons behind the differences in the timescale of regularity learning found in this thesis?

One possibility is that predictive processing underlying the results of Chapter 2 compared to Chapters 3 and 4 used different prior knowledge acquired from regularity learning. The regularities in the environment between chapters may have been extracted by different predictive processing systems working at different time scales. This notion is compatible with the taxonomy of sensory expectation proposed by Seriès and Seitz, who suggest two different categories of sensory expectation. (Seriès & Seitz, 2013). Seriès and Seitz described the two types of expectation as contextual and structural expectations (Figure 5-1):

"Structural expectations are the 'default' expectations that human observers use based on implicit learning of the statistics of the natural environment. These expectations usually reflect long-term learning over the lifetime, or may be innate… Contextual expectations, on the other hand, can be manipulated rapidly, explicitly or implicitly, through sensory, or by the spatial, temporal, or stimulus context in which a stimulus is shown" (Seriès & Seitz, 2013, p. 2).

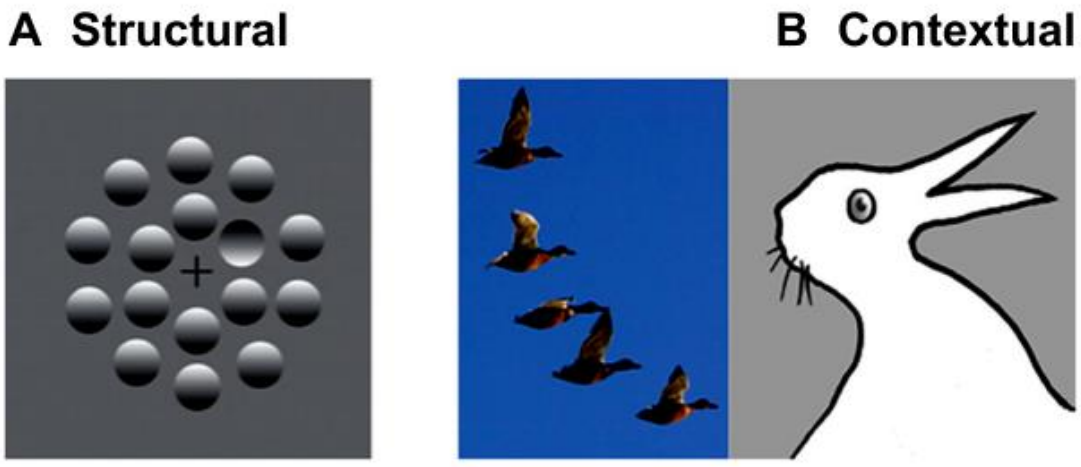

Figure 5-1 Structural vs. contextual expectations. (A) Our priors tells us that light must come from the top of the figure, known as the "light-from-above" prior. Therefore, this percept is biased by a long-term prior knowledge of the light direction (structural expectation). (B) This ambiguous and bistable figure was more frequently perceived as a duck, rather than the alternative (a rabbit) due being paired with the presentation of a prime of flying ducks. This biased percept shows the short-term influence from the context (context expectation). (figure adapted from Seriès & Seitz, 2013).

Following from this perspective, the associative learning used in the MIB study in Chapter 2 may have caused structural expectations to develop, which were established through extensive training between the auditory cue and the colour of the target. Through the intensive exposure of the contingency between auditory and visual stimuli, the neural system may have built firm long-term knowledge of the cross-modal regularity. In contrast the sequence and duration repetitions in Chapter 3 and 4 may have formed a temporary context across stimuli leading to a short-term, flexible contextual expectation. The contextual expectation quickly adapts to the new context, as shown by the rapid enhancement of the echo response to a new stimulus sequence found in Chapter 3. The new sequence immediately re-established the linear enhancement of echo response. Similarly, the results of Chapter 4 demonstrate that the MMN can quickly adapt to new stimulus features. A deviant duration can immediately change the neural response elicited by the following standard duration, suggesting that at least for the MMN neural systems can very rapidly form contextual expectations. Note that, although the associative training described in Chapter 2 is extensive, whether the training is able to establish a firm structural expectation (e.g., the light from above effect) is still unknown. Future research is needed to clarify how extensive training influences structural expectations.

Together the results of this thesis show that the time course of regularity learning varied across experiments. This may have been due to the differing structural and contextual expectations developed by each experimental manipulation. I suggest that the change of conscious access speed

found in Chapter 2 may be affected by structural expectations that are developed by a large amount of learning, such as the associative training used in the MIB experiment in Chapter 2.

5.4 The neural signatures of predictive processing may not always be dissociable

Previous neurophysiological evidence of predictive processing has mainly focused on the neural correlates of the prediction error (See review in Bastos et al., 2012; Karl Friston, 2005). That is, if neural responses are enhanced when predictions are inconsistent with sensory inputs, the activities are considered to represent prediction errors. Evidence for this claim has been found in human EEG and MEG (Baldeweg, 2006; Khouri & Nelken, 2015; Stefanics et al., 2014, p. 2; István Winkler & Czigler, 2012), human fMRI (Alink, Schwiedrzik, Kohler, Singer, & Muckli, 2010; Kok & de Lange, 2014; Summerfield & Egner, 2009), and also at the neuronal level in an animal study (Keller, Bonhoeffer, & Hübener, 2012). Consistent with these findings, Chapter 4 showed that when sensory events were presented at unexpected timing (the deviant duration) the EEG signal showed a large deflection, which I suggest represents a neural signature of the prediction error.

However, Chapter 3 shows that the perceptual echo, a potential neurophysiological index of low-level visual predictive mechanism, was *enhanced* by repetitions of a stimulus sequence. Furthermore, the enhancement was not associated with a power change in the raw EEG alpha oscillation. As discussed in Section 3.4, this result suggests that the activation level of raw neural responses (raw EEG alpha) in this low-level visual predictive processing were *not* modulated by the predictability of sensory inputs. The enhancement of perceptual echo by sequence repetitions was more likely caused by the increasing similarity between sequence specific information (luminance) and the evoked EEG response (see 3.4). As mentioned in the Chapter 3, this effect can be thought of as sharpening of the neural assemblies encoding the visual sequences in order to form

a more precise representation of the sequence. The central assumptions of the model created in Chapter 3 also suggests that the simulated EEG signals was modulated not only by prediction error computation but also by neural activity representing the prediction of luminance values.

Taken together, the results of Chapter 3 suggest that the prediction and prediction error neurons may be spatially indistinguishable using a spatially coarse neural imaging method, such as EEG. Previous research has shown that neurons responsible for prediction and prediction error computations can be found within the same cortical areas. For example, in a mouse study, Keller, Bonhoeffer, and Hübener (2012) investigated predictive neural signals from motor areas and prediction error signals in the primary visual cortex by manipulating visual-flow feedback during locomotion in a virtual reality environment using two-photon imaging of neurons expressing a genetically encoded calcium indicator. The results showed that neural activates in mouse primary visual cortex were driven by locomotion and by the mismatch between actual and expected visual feedback, suggesting that mouse early visual areas are involved with both predictions and prediction error computations. More importantly, the result showed no indications for spatial clustering of different response types, confirming that different neural indexes of predictive processing would not be dissociable using scalp recorded EEG.

In summary, the results of this thesis related to the perceptual echo demonstrates that the neural evidence of predictive processing may not always be revealed by prediction error activity alone. In the case of the perceptual echo I did not observe an increase in neural activity (alpha oscillation) as would be expected with a prediction error when a new sequence was presented. Therefore, it is possible to observe neural indexes of predictive processing without an explicit prediction error signal.

Future research requires more sophisticated methods of data analysis to validate the predictive processing framework across multiple experiments. Promising methods have already been applied

to this aim. For example, Garrido et al. showed that MMN, an index of prediction error, could be explained by changes in the strength of the functional connectivity between cortical sources (Garrido et al., 2008). Additionally, Arnal et al. found, using MEG, that delta neural oscillations (3–4 Hz) conveyed prediction signals from higher-order speech areas, while low-beta (14–15 Hz) / high-gamma (60–80 Hz) coupling represented prediction errors within multisensory areas (Arnal et al., 2011; see also Sedley et al., 2016).

5.5 Limitations and future research

5.5.1 How does the method of acquisition of predictive information affect predictive processing?

Participants in all the studies within this thesis acquired predictive information from exposure to environmental regularities manipulated experimentally. This thesis focused on experimental manipulations which meant that predictive information and the statistical regularities had to be spontaneously extracted and learned from the experimental environments. In Chapter 2, participants learned statistical relationships between auditory cues and visual targets from intensive associative learning training. In Chapter 3, participants were not informed about the repetitions of luminance sequences and when probed about their experiences in the study reported that they were unaware of any repetitions throughout the whole experiments. In Chapter 4, participants were not informed about the statistical structure of standard and deviant trials. However, in real life, predictive information is not always learned through the statistical structure of external events. Instead, humans are able to learn and update prior knowledge and predictive relationships by other means, for example verbal instruction. To my knowledge, no study has been conducted to systematically investigate the difference of using prior knowledge learned from exposure to statistical regularities compared to explicit linguistic information about the same predictive

relationship.

Nevertheless, some behavioural and neural evidence suggests that neural systems processes predictive information acquired through different routes in dissimilar ways. For example, linguistic rules when learning a second language (L2) can be acquired either by exposure to form implicit knowledge or by overtly being taught the form explicit knowledge (Ellis, 2005). Hulstijn (2002) suggests that implicit and explicit L2 knowledge involve different acquisition mechanisms that are represented in different brain areas (Paradis, 1994), and the different types of L2 knowledge need to be accessed by different tasks (Ellis, 2005). Additional support for the idea that predictive information may be processed in differing ways comes from imaging studies that find that implicit and explicit predictive knowledge induces different neural responses. In a fMRI study by Aizenstein et al. participants viewed consecutive sequences consisting of a series of items with two attributes, colour and location (Aizenstein et al., 2004). Participants were required to press different buttons according to the colour or shape of an item. The colours and shapes of items were determined independently and follow two different first-order Markov chains. Participants were only informed about the existence of the statistical relationship of one stimulus feature and responded to it as quickly as possible (task-relevant) and learned the sequence rule of the other feature (task-irrelevant) implicitly through exposure. The results showed that when the task-relevant feature was consistent with predictions that were generated from explicit knowledge, visual regions (V1, V2, and V3) showed increased activity. Interestingly, when the task-irrelevant features were consistent with predictions from implicit knowledge, the same visual regions showed decreased activity. These results suggest that different neural mechanisms were involved in predictive processing when using explicit and implicit knowledge.

Taken together, future behavioural and imaging studies should consider the impact of explicitly and implicitly learned knowledge in the predictive processing framework. More research

is needed to systematically examine how different prior knowledge influences the neural computation of prediction and how the route of the acquisition of statistical regularities affects how the brain processes this information.

5.5.2 Can sequence knowledge become conscious?

Chapter 3 showed that sequence repetitions linearly enhanced perceptual echo amplitude in the absence of awareness of the sequence repetitions. This suggests that the human visual system encodes knowledge of dynamic luminance sequences unconsciously, even when a sequence was repeated up to 6 times. However, the current study did not test whether this implicit knowledge may become explicit with many more repetitions of a sequence.

Previous research has demonstrated that humans are capable of learning implicit sequence knowledge underlying motor, cognitive and even social skills (Cleeremans, Destrebecqz, & Boyer, 1998; Kaufman et al., 2010; Lieberman, 2000; Nemeth et al., 2011; Romano Bergstrom, Howard, & Howard, 2012). However, some evidence suggests, after repeated exposures of a stimulus sequence, human subjects may be able to acquire explicit knowledge of the sequence presentation rules. A classic example of this phenomenon can be found in the serial reaction time task (Cleeremans & McClelland, 1991; Knopman & Nissen, 1987; Nissen & Bullemer, 1987; Peigneux et al., 2000, 2003; Willingham, Nissen, & Bullemer, 1989). In this task, subjects are required to respond as fast and as accurately as possible to the appearance of a target stimulus arranged horizontally on a screen by pressing the spatially corresponding key. Without informing the participants, the order of target locations used in these studies follows deterministic, probabilistic, or grammatical rules. After a short training period on these tasks, participants generally improve their reaction times without being able to explicitly describe the rules that determine the stimulus

order, suggesting that they have acquired implicit knowledge about the rules. Interestingly, individuals often gain explicit awareness of the sequence in the latter phases of the practice session (Nissen & Bullemer, 1987), suggesting that the explicit knowledge system takes more time and exposure to acquire the embedded rules of target location.

The current experimental design of Chapter 3 was not optimised to assess the impact of sequence repetition on conscious awareness of these repetitions. Future research to investigate this research question should use a behavioural recognition task. For example, Agus et al. investigated sequence learning in humans within the auditory domain and demonstrated that the auditory system can rapidly learn and form memories of repeated sequences constructed of Gaussian random noise (Agus et al., 2010). In this study, participants were instructed to perform a judgment task to report whether or not they heard a repetition of an audio clip. The result showed that the learning of auditory sequences occurred very rapidly, participants were able to learn a novel sequence in less than 10 repetitions. Similarly, previous investigation on artificial grammar learning (Reber, 1967) used a binary judgment task to assess conscious knowledge of embedded sequence rules (Dienes & Altmann, 1997; Zoltán Dienes & Scott, 2005). Participants are first exposed to sequences generated by one grammars (grammar A) in the training phase. In the test phase, sequences generated by grammar A and another grammar (grammar B) are randomly presented. Participants are asked to classify the test items (A or B). Using the binary judgment task, the basic conscious knowledge of artificial grammars embedded in sequences can be examined. A similar experimental design could be adopted to evaluate whether human participants were able to acquire explicit knowledge of the fast changing dynamic luminance sequences used in Chapter 3.

Chapter 6
Conclusion

This thesis investigated the behavioural and neural correlates of predictive processing and regularity learning across differing levels of predictive information and conscious access. It has demonstrated how predictive influences on conscious access develop across time via training, and how they permeate and shape conscious experiences more deeply and broadly than previously thought. It provides electrophysiological and modelling evidence that the perceptual echo reflects the existence of a previously undiscovered temporal regularity learning and predictive processing mechanism. It identified a neural correlate of temporal prediction based on duration processing that was shared between visual and auditory sensory predictive processing.

Overall this thesis has added to our understanding of the relationship between unconscious statistical inference and conscious access by demonstrating that predictive information is processed unconsciously but can influence conscious access. This thesis has also shown that predictive processing operates across differing levels of sensory information, low-level, cross-modal, and supra-modal. Finally, this thesis suggests that predictive processing mechanisms use different types of regularity learning to generate valid predictions about future events.

Together this thesis has added to our understanding of predictive processing and regularity learning across differing types of predictive information.

References

Agus, T. R., Thorpe, S. J., & Pressnitzer, D. (2010). Rapid Formation of Robust Auditory Memories: Insights from Noise. *Neuron*, *66*(4), 610–618. https://doi.org/10.1016/j.neuron.2010.04.014

Aizenstein, H. J., Stenger, V. A., Cochran, J., Clark, K., Johnson, M., Nebes, R. D., & Carter, C. S. (2004). Regional Brain Activation during Concurrent Implicit and Explicit Sequence Learning. *Cerebral Cortex*, *14*(2), 199–208. https://doi.org/10.1093/cercor/bhg119

Alais, D., & Burr, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, *14*(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029

Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus Predictability Reduces Responses in Primary Visual Cortex. *The Journal of Neuroscience*, *30*(8), 2960–2966. https://doi.org/10.1523/JNEUROSCI.3730-10.2010

Alsius, A., & Munhall, K. G. (2013). Detection of Audiovisual Speech Correspondences Without Visual Awareness. *Psychological Science*, *24*(4), 423–431. https://doi.org/10.1177/0956797612457378

Altieri, N. (2014). Multisensory integration, learning, and the predictive coding hypothesis. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00257

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, *96*(4), 703.

Arnal, L. H., Wyart, V., & Giraud, A.-L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nature Neuroscience*, *14*(6), 797–801. https://doi.org/10.1038/nn.2810

Baldeweg, T. (2006). Repetition effects to sounds: evidence for predictive coding in the auditory system. *Trends in Cognitive Sciences*, *10*(3), 93–94. https://doi.org/10.1016/j.tics.2006.01.010

Baldeweg, T. (2007). ERP Repetition Effects and Mismatch Negativity Generation. *Journal of Psychophysiology*, *21*(3), 204–213. https://doi.org/10.1027/0269-8803.21.34.204

Baldeweg, T., Klugman, A., Gruzelier, J., & Hirsch, S. R. (2004). Mismatch negativity potentials and cognitive impairment in schizophrenia. *Schizophrenia Research*, *69*(2–3), 203–217. https://doi.org/10.1016/j.schres.2003.09.009

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. (2012). Canonical Microcircuits for Predictive Coding. *Neuron*, *76*(4), 695–711. https://doi.org/10.1016/j.neuron.2012.10.038

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *JOSA A*, *20*(7), 1391–1397.

Bendixen, A., Schröger, E., & Winkler, I. (2009). I Heard That Coming: Event-Related Potential Evidence for Stimulus-Driven Prediction in the Auditory System. *Journal of Neuroscience*, *29*(26), 8447–8451. https://doi.org/10.1523/JNEUROSCI.1493-09.2009

Bi, G.-Q., & Poo, M.-M. (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, *18*(24), 10464–10472.

Bonneh, Y., Cooperman, A., & Sagi, D. (2001). Motion-induced blindness in normal observers. *Nature*, *411*(6839), 798–801. https://doi.org/10.1038/35081073

Bonneh, Y. S., Donner, T. H., Sagi, D., Fried, M., Cooperman, A., Heeger, D. J., & Arieli, A. (2010). Motion-induced blindness and microsaccades: Cause and effect. *Journal of Vision*, *10*(14), 1–15. https://doi.org/10.1167/10.14.22

Bracewell, R. (2004). *Fourier analysis and imaging*. Springer Science & Business Media.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brainard, D. H., & Freeman, W. T. (1997). Bayesian color constancy. *Journal of the Optical Society of America A*, *14*(7), 1393–1411. https://doi.org/10.1364/JOSAA.14.001393

Büchel, C., Coull, J. T., & Friston, K. J. (1999). The Predictive Value of Changes in Effective Connectivity for Human Learning. *Science*, *283*(5407), 1538–1541. https://doi.org/10.1126/science.283.5407.1538

Buckner, R. L., Goodman, J., Burock, M., Rotte, M., Koutstaal, W., Schacter, D., … Dale, A. M. (1998). Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI. *Neuron*, *20*(2), 285–296.

Bueti, D., Bahrami, B., & Walsh, V. (2008). Sensory and Association Cortex in Time Perception. *Journal of Cognitive Neuroscience*, *20*(6), 1054–1062. https://doi.org/10.1162/jocn.2008.20060

Buonomano, D. V., & Karmarkar, U. R. (2002). Book Review: How Do We Tell Time? *The Neuroscientist*, *8*(1), 42–51.

Canales-Johnson, A., Silva, C., Huepe, D., Rivera-Rei, Á., Noreika, V., Garcia, M. del C., … Bekinschtein, T. A. (2015). Auditory Feedback Differentially Modulates Behavioral and Neural Markers of Objective and Subjective Performance When Tapping to Your Heartbeat. *Cerebral Cortex*, bhv076. https://doi.org/10.1093/cercor/bhv076

Chang, C., & Lin, C.-J. (2001). *LIBSVM: a Library for Support Vector Machines*.

Chen, Y., Huang, X., Luo, Y., Peng, C., & Liu, C. (2010). Differences in the neural basis of automatic auditory and visual time perception: ERP evidence from an across-modal delayed response oddball task. *Brain Research*, *1325*, 100–111. https://doi.org/10.1016/j.brainres.2010.02.040

Chen, Y.-C., Yeh, S.-L., & Spence, C. (2011). Crossmodal constraints on human perceptual awareness: auditory semantic modulation of binocular rivalry. *Frontiers in Psychology*, *2*, 212. https://doi.org/10.3389/fpsyg.2011.00212

Clark, A. (2012). Whatever Next ? Predictive Brains , Situated Agents , and the Future of Cognitive Science . *Behavioral and Brain Sciences*, 1–86.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: news from the front. *Trends in Cognitive Sciences*, *2*(10), 406–416. https://doi.org/10.1016/S1364-6613(98)01232-7

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, *120*(3), 235–253. https://doi.org/10.1037/0096-3445.120.3.235

Conrad, V., Bartels, A., Kleiner, M., & Noppeney, U. (2010). Audiovisual interactions in binocular rivalry. *Journal of Vision*, *10*(10). https://doi.org/10.1167/10.10.27

Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences*, *5*(12), 539–546. https://doi.org/10.1016/S1364-6613(00)01800-3

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clinical Neurophysiology*, *115*(4), 732–744. https://doi.org/10.1016/j.clinph.2003.11.021

Davis, B., & Hasson, U. (2016). Predictability of what or where reduces brain activity, but a bottleneck occurs when both are predictable. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2016.06.001

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz machine. *Neural Computation*, *7*(5), 889–904.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

den Ouden, H. E. M., Friston, K., Daw, N. D., McIntosh, A. R., & Stephan, K. E. (2009). A Dual Role for Prediction Error in Associative Learning. *Cerebral Cortex*, *19*(5), 1175–1185. https://doi.org/10.1093/cercor/bhn161

Deneve, S., & Pouget, A. (2004). Bayesian multisensory integration and cross-modal spatial links. *Journal of Physiology-Paris*, *98*(1–3), 249–258. https://doi.org/10.1016/j.jphysparis.2004.03.011

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences*, *93*(24), 13494–13499.

Dienes, Z. (2011). Bayesian Versus Orthodox Statistics: Which Side Are You On? *Perspectives on Psychological Science*, *6*(3), 274–290. https://doi.org/10.1177/1745691611406920

Dienes, Z., & Altmann, G. (1997). Transfer of implicit knowledge across domains: How implicit and how abstract. *How Implicit Is Implicit Learning*, *5*, 107–123.

Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: distinguishing structural knowledge and judgment knowledge. *Psychological Research*, *69*(5–6), 338–351. https://doi.org/10.1007/s00426-004-0208-3

Droit-Volet, S., Meck, W. H., & Penney, T. B. (2007). Sensory modality and time perception in children and adults. *Behavioural Processes*, *74*(2), 244–250. https://doi.org/10.1016/j.beproc.2006.09.012

Durrant, S. J., Taylor, C., Cairney, S., & Lewis, P. A. (2011). Sleep-dependent consolidation of statistical learning. *Neuropsychologia*, *49*(5), 1322–1331. https://doi.org/10.1016/j.neuropsychologia.2011.02.015

Eagleman, S. L., & Dragoi, V. (2012). Image sequence reactivation in awake V4 networks. *Proceedings of the National Academy of Sciences*, *109*(47), 19450–19455. https://doi.org/10.1073/pnas.1212059109

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27*(02), 141–172.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. https://doi.org/10.1038/415429a

Felleman, D. J., & Essen, D. C. V. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex*, *1*(1), 1–47. https://doi.org/10.1093/cercor/1.1.1

Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature*, *425*(6958), 614–616. https://doi.org/10.1038/nature01951

Fischer, C., Morlet, D., Bouchet, P., Luaute, J., Jourdan, C., & Salord, F. (1999). Mismatch negativity and late auditory evoked potentials in comatose patients. *Clinical Neurophysiology*, *110*(9), 1601–1610. https://doi.org/10.1016/S1388-2457(99)00131-5

Fischer, C., Morlet, D., & Giard, M. (2000). Mismatch negativity and N100 in comatose patients. *Audiology & Neuro-Otology*, *5*(3–4), 192–197. https://doi.org/13880

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends in Cognitive Sciences*, *14*(3), 119–130.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, *45*(1), 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x

Ford, J. M., & Hillyard, S. A. (1981). Event-related potentials (ERPs) to interruptions of a steady rhythm. *Psychophysiology*, *18*(3), 322–330.

Friston, K. (2002). Functional integration and inference in the brain. *Progress in Neurobiology*, *68*(2), 113–143. https://doi.org/10.1016/S0301-0082(02)00076-X

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K. (2012). The history of the future of the Bayesian brain. *NeuroImage*, *62*(2), 1230–1233. https://doi.org/10.1016/j.neuroimage.2011.10.004

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211–1221. https://doi.org/10.1098/rstb.2008.0300

Friston, K., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, *159*(3), 417–458. https://doi.org/10.1007/s11229-007-9237-y

Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal Predictive Codes for Spoken Words in Auditory Cortex. *Current Biology*, *22*(7), 615–621. https://doi.org/10.1016/j.cub.2012.02.015

Garrido, M. I., Friston, K., Kiebel, S. J., Stephan, K. E., Baldeweg, T., & Kilner, J. M. (2008). The functional anatomy of the MMN: A DCM study of the roving paradigm. *NeuroImage*, *42*(2), 936–944. https://doi.org/10.1016/j.neuroimage.2008.05.018

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, *120*(3), 453–463. https://doi.org/10.1016/j.clinph.2008.11.029

Gavornik, J. P., & Bear, M. F. (2014). Higher brain functions served by the lowly rodent primary visual cortex. *Learning & Memory*, *21*(10), 527–533. https://doi.org/10.1101/lm.034355.114

Gheysen, F., Van Opstal, F., Roggeman, C., Van Waelvelde, H., & Fias, W. (2011). The Neural Basis of Implicit Perceptual Sequence Learning. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00137

Goldstein, M. H., Waterfall, H. R., Lotem, A., Halpern, J. Y., Schwade, J. A., Onnis, L., & Edelman, S. (2010). General cognitive principles for learning structure in time and space. *Trends in Cognitive Sciences*, *14*(6), 249–258. https://doi.org/10.1016/j.tics.2010.02.004

Gregory, R. L. (1968). Perceptual Illusions and Brain Models. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *171*(1024), 279–296.

Gregory, R. L. (1980). Perceptions as Hypotheses. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *290*(1038), 181–197. https://doi.org/10.1098/rstb.1980.0090

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*(1), 14–23. https://doi.org/10.1016/j.tics.2005.11.006

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, *24*(1), 187–203.

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychologica*, *107*(1), 293–321.

Grimm, S., Roeber, U., Trujillo-Barreto, N. J., & Schröger, E. (2006). Mechanisms for detecting auditory temporal and spectral deviations operate over similar time windows but are divided

differently between the two hemispheres. *NeuroImage*, *32*(1), 275–282. https://doi.org/10.1016/j.neuroimage.2006.03.032

Grimm, S., Widmann, A., & Schröger, E. (2004). Differential processing of duration changes within short and long sounds in humans. *Neuroscience Letters*, *356*(2), 83–86. https://doi.org/10.1016/j.neulet.2003.11.035

Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics*, *72*(3), 561–582. https://doi.org/10.3758/APP.72.3.561

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review: Mass univariate analysis of ERPs/ERFs I: Review. *Psychophysiology*, *48*(12), 1711–1725. https://doi.org/10.1111/j.1469-8986.2011.01273.x

Grube, M., Cooper, F. E., Chinnery, P. F., & Griffiths, T. D. (2010). Dissociation of duration-based and beat-based auditory timing in cerebellar degeneration. *Proceedings of the National Academy of Sciences*, *107*(25), 11597–11601. https://doi.org/10.1073/pnas.0910473107

Grube, M., Lee, K.-H., Griffiths, T. D., Barker, A. T., & Woodruff, P. W. (2010). Transcranial magnetic theta-burst stimulation of the human cerebellum distinguishes absolute, duration-based from relative, beat-based perception of subsecond time intervals. *Auditory Cognitive Neuroscience*, *1*, 171. https://doi.org/10.3389/fpsyg.2010.00171

Guyonneau, R., VanRullen, R., & Thorpe, S. J. (2005). Neurons Tune to the Earliest Spikes Through STDP. *Neural Computation*, *17*(4), 859–879. https://doi.org/10.1162/0899766053429390

Haenschel, C. (2005). Event-Related Brain Potential Correlates of Human Auditory Sensory Memory-Trace Formation. *Journal of Neuroscience*, *25*(45), 10494–10501. https://doi.org/10.1523/JNEUROSCI.1227-05.2005

Hamm, A. O., & Vaitl, D. (1996). Affective learning: Awareness and aversion. *Psychophysiology*, *33*(6), 698–710. https://doi.org/10.1111/j.1469-8986.1996.tb02366.x

Hatfield, G. (2005). Perception as Unconscious Inference. In D. H. P. D. essor of Psychology & R. of C. Science (Eds.), *Perception and the Physical World* (pp. 113–143). John Wiley & Sons, Ltd. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/0470013427.ch5/summary

Helmholtz, H. von. (1866). Concerning the perceptions in general. *Treatise on Physiological Optics*, *3*.

Helmholtz, H. von. (1867). *Treatise on Physiological Optics*, *3*.

Henson, R., Shallice, T., & Dolan, R. (2000). Neuroimaging evidence for dissociable forms of repetition priming. *Science*, *287*(5456), 1269–1272.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. *Neuron*, *69*(3), 407–422. https://doi.org/10.1016/j.neuron.2011.01.019

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, *4*(12). https://doi.org/10.1167/4.12.1

Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in Cognitive Sciences*, *11*(10), 428–434. https://doi.org/10.1016/j.tics.2007.09.004

Hock, H. S., Scott, J., & Schöner, G. (1993). Bistability and hysteresis in the organization of apparent motion patterns. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(1), 63–80. https://doi.org/10.1037/0096-1523.19.1.63

Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology*, *3*(April), 96. https://doi.org/10.3389/fpsyg.2012.00096

Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press.

Hohwy, J. (2014). *The Predictive Mind*. Oxford, United Kingdom ; New York, NY, United States of America: Oxford University Press.

Horn, B. K. P., & Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, *17*(1), 185–203. https://doi.org/10.1016/0004-3702(81)90024-2

Horváth, J., Czigler, I., Jacobsen, T., Maess, B., Schröger, E., & Winkler, I. (2008). MMN or no MMN: No magnitude of deviance effect on the MMN amplitude. *Psychophysiology*, *45*(1), 60–69. https://doi.org/10.1111/j.1469-8986.2007.00599.x

Horváth, J., Roeber, U., & Schröger, E. (2009). The utility of brief, spectrally rich, dynamic sounds in the passive oddball paradigm. *Neuroscience Letters*, *461*(3), 262–265.

Hruby, T., & Marsalek, P. (2002). Event-related potentials-the P3 wave. *Acta Neurobiologiae Experimentalis*, *63*(1), 55–63.

Hsieh, P.-J., & Tse, P. U. (2009). Microsaccade Rate Varies with Subjective Visibility during Motion-Induced Blindness. *PLoS ONE*, *4*(4), e5163. https://doi.org/10.1371/journal.pone.0005163

Hsu, W.-Y., Cheng, C.-H., Lin, H.-C., Liao, K.-K., Wu, Z.-A., Ho, L.-T., & Lin, Y.-Y. (2010). Memory-based mismatch response to changes in duration of auditory stimuli: An MEG study. *Clinical Neurophysiology*, *121*(10), 1744–1750. https://doi.org/10.1016/j.clinph.2010.04.003

Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), 580–593. https://doi.org/10.1002/wcs.142

Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, *28*(2), 229–289.

Huettel, S. A., Song, A. W., & McCarthy, G. (2005). Decisions under uncertainty: probabilistic context influences activation of prefrontal and parietal cortices. *The Journal of Neuroscience*, *25*(13), 3304–3311.

Hürlimann, F., Kiper, D. C., & Carandini, M. (2002). Testing the Bayesian model of perceived speed. *Vision Research*, *42*(19), 2253–2257.

Ivry, R. B., & Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception. *Trends in Cognitive Sciences*, *12*(7), 273–280. https://doi.org/10.1016/j.tics.2008.04.002

Jääskeläinen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levänen, S., … Belliveau, J. W. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(17), 6809–6814. https://doi.org/10.1073/pnas.0303760101

Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, *39*(21), 3621–3629.

Jacobsen, T., & Schröger, E. (2003). Measuring duration mismatch negativity. *Clinical Neurophysiology*, *114*(6), 1133–1143. https://doi.org/10.1016/S1388-2457(03)00043-9

JASP Team. (2016). JASP (Version 0.7.5.5)[Computer software].

Joutsiniemi, S.-L., Ilvonen, T., Sinkkonen, J., Huotilainen, M., Tervaniemi, M., Lehtokoski, A., … Näätänen, R. (1998). The mismatch negativity for duration decrement of auditory stimuli in healthy subjects. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *108*(2), 154–159. https://doi.org/10.1016/S0168-5597(97)00082-8

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, *82*(1), 35–45. https://doi.org/10.1115/1.3662552

Kanai, R., & Verstraten, F. A. J. (2005). Perceptual manifestations of fast neural plasticity: Motion priming, rapid motion aftereffect and perceptual sensitization. *Vision Research*, *45*(25–26), 3109–3116. https://doi.org/10.1016/j.visres.2005.05.014

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences*, *88*(11), 4966–4970.

Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J., & Sagi, D. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill. *Science*, *265*(5172), 679–682. https://doi.org/10.1126/science.8036518

Katz, J. (2014). *Handbook of Clinical Audiology* (North American ed edition). Philadelphia: Lippincott Williams and Wilkins.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*(3), 321–340.

Keller, G. B., Bonhoeffer, T., & Hübener, M. (2012). Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. *Neuron*, *74*(5), 809–15. https://doi.org/10.1016/j.neuron.2012.03.040

Kentridge, R. W., Heywood, C. A., & Weiskrantz, L. (2004). Spatial attention speeds discrimination without awareness in blindsight. *Neuropsychologia*, *42*(6), 831–835. https://doi.org/10.1016/j.neuropsychologia.2003.11.001

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object Perception as Bayesian Inference. *Annual Review of Psychology*, *55*(1), 271–304. https://doi.org/10.1146/annurev.psych.55.090902.142005

Kersten, D., & Schrater, P. (2002). *Pattern inference theory: A probabilistic approach to vision*. na.

Kersten, D., & Yuille, A. (2003). Bayesian models of object perception. *Current Opinion in Neurobiology*, *13*(2), 150–158. https://doi.org/10.1016/S0959-4388(03)00042-4

Khouri, L., & Nelken, I. (2015). Detecting the unexpected. *Current Opinion in Neurobiology*, *35*, 142–147. https://doi.org/10.1016/j.conb.2015.08.003

King, A. J., & Palmer, A. R. (1985). Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. *Experimental Brain Research*, *60*(3), 492–500. https://doi.org/10.1007/BF00236934

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. https://doi.org/10.1016/j.tics.2014.01.002

King, J.-R., Gramfort, A., Schurger, A., Naccache, L., & Dehaene, S. (2014). Two Distinct Dynamic Modes Subtend the Detection of Unexpected Sounds. *PLoS ONE*, *9*(1), e85791. https://doi.org/10.1371/journal.pone.0085791

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? In *30th European Conference on Visual Perception*.

Kleinschmidt, A., Büchel, C., Hutton, C., Friston, K., & Frackowiak, R. S. . (2002). The Neural Structures Expressing Perceptual Hysteresis in Visual Letter Recognition. *Neuron*, *34*(4), 659–666. https://doi.org/10.1016/S0896-6273(02)00694-3

Knight, D. C., Nguyen, H. T., & Bandettini, P. A. (2003). Expression of conditional fear with and without awareness. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(25), 15280–15283. https://doi.org/10.1073/pnas.2535780100

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. https://doi.org/10.1016/j.tins.2004.10.007

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian Inference*. Cambridge University Press.

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, *43*(24), 2539–2558. https://doi.org/10.1016/S0042-6989(03)00458-9

Knopman, D. S., & Nissen, M. J. (1987). Implicit learning in patients with probable Alzheimer's disease. *Neurology*, *37*(5), 784–784.

Kok, P., Brouwer, G. J., Gerven, M. A. J. van, & de Lange, F. P. (2013). Prior Expectations Bias Sensory Representations in Visual Cortex. *The Journal of Neuroscience*, *33*(41), 16275–16284. https://doi.org/10.1523/JNEUROSCI.0742-13.2013

Kok, P., & de Lange, F. P. (2014). Shape Perception Simultaneously Up- and Downregulates Neural Activity in the Primary Visual Cortex. *Current Biology*, *24*(13), 1531–1535. https://doi.org/10.1016/j.cub.2014.05.042

Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior Expectations Evoke Stimulus Templates in the Primary Visual Cortex. *Journal of Cognitive Neuroscience*, 1–9. https://doi.org/10.1162/jocn_a_00562

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron*, *75*(2), 265–270. https://doi.org/10.1016/j.neuron.2012.04.034

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal Inference in Multisensory Perception. *PLoS ONE*, *2*(9), e943. https://doi.org/10.1371/journal.pone.0000943

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247. https://doi.org/10.1038/nature02169

Kourtzi, Z., & Kanwisher, N. (2001). Representation of Perceived Object Shape by the Human Lateral Occipital Complex. *Science*, *293*(5534), 1506–1509. https://doi.org/10.1126/science.1061133

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, *20*(7), 1434–1448.

Li, N., & DiCarlo, J. J. (2008). Unsupervised Natural Experience Rapidly Alters Invariant Object Representation in Visual Cortex. *Science*, *321*(5895), 1502–1507. https://doi.org/10.1126/science.1160028

Lieberman, M. D. (2000). Intuition: a social cognitive neuroscience approach. *Psychological Bulletin*, *126*(1), 109.

Liégeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P., & Chauvel, P. (1994). Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *92*(3), 204–214. https://doi.org/10.1016/0168-5597(94)90064-7

Loschky, L., McConkie, G., Yang, J., & Miller, M. (2005). The limits of visual resolution in natural scene viewing. *Visual Cognition*, *12*(6), 1057–1092. https://doi.org/10.1080/13506280444000652

Loveless, N., Levänen, S., Jousmäki, V., Sams, M., & Hari, R. (1996). Temporal integration in auditory sensory memory: neuromagnetic evidence. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *100*(3), 220–228. https://doi.org/10.1016/0168-5597(95)00271-5

Luft, C. D. B., Meeson, A., Welchman, A. E., & Kourtzi, Z. (2015). Decoding the future from past experience: learning shapes predictions in early visual cortex. *Journal of Neurophysiology*, *113*(9), 3159–3171. https://doi.org/10.1152/jn.00753.2014

Lunghi, C., Binda, P., & Morrone, M. C. (2010). Touch disambiguates rivalrous perception at early stages of visual analysis. *Current Biology*, *20*(4), R143–R144. https://doi.org/10.1016/j.cub.2009.12.015

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1303312110

Lustig, C., & Meck, W. H. (2011). Modality differences in timing and temporal memory throughout the lifespan. *Brain and Cognition*, *77*(2), 298–303. https://doi.org/10.1016/j.bandc.2011.07.007

MacKay, D. M. (1956). *The epistemological problem for automata*.

Maimon, O., & Rokach, L. (2005). *Data mining and knowledge discovery handbook* (Vol. 2). Springer.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Markram, H., Lübke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of Synaptic Efficacy by Coincidence of Postsynaptic APs and EPSPs. *Science*, *275*(5297), 213–215. https://doi.org/10.1126/science.275.5297.213

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. *Henry Holt and Co., New York*.

Martinez-Conde, S., Macknik, S. L., Troncoso, X. G., & Dyar, T. A. (2006). Microsaccades Counteract Visual Fading during Fixation. *Neuron*, *49*(2), 297–305. https://doi.org/10.1016/j.neuron.2005.11.033

Masquelier, T., Guyonneau, R., & Thorpe, S. J. (2008a). Competitive STDP-Based Spike Pattern Learning. *Neural Computation*, *21*(5), 1259–1276. https://doi.org/10.1162/neco.2008.06-08-804

Masquelier, T., Guyonneau, R., & Thorpe, S. J. (2008b). Spike Timing Dependent Plasticity Finds the Start of Repeating Patterns in Continuous Spike Trains. *PLOS ONE*, *3*(1), e1377. https://doi.org/10.1371/journal.pone.0001377

Mauk, M. D., & Buonomano, D. V. (2004). The Neural Basis of Temporal Processing. *Annual Review of Neuroscience*, *27*(1), 307–340. https://doi.org/10.1146/annurev.neuro.27.070203.144247

Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, *29*(6), 317–322. https://doi.org/10.1016/j.tins.2006.04.001

Mcintosh, A. R., & Gonzalez-Lima, F. (1998). Large-Scale Functional Connectivity in Associative Learning: Interrelations of the Rat Auditory, Visual, and Limbic Systems. *Journal of Neurophysiology*, *80*(6), 3148–3162.

Melloni, L., Schwiedrzik, C. M., Müller, N., Rodriguez, E., & Singer, W. (2011). Expectations Change the Signatures and Timing of Electrophysiological Correlates of Perceptual Awareness. *The Journal of Neuroscience*, *31*(4), 1386–1396. https://doi.org/10.1523/JNEUROSCI.4570-10.2011

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences*, *108*(48), 19401–19406. https://doi.org/10.1073/pnas.1112895108

Mitroff, S. R., & Scholl, B. J. (2005). Forming and updating object representations without awareness: evidence from motion-induced blindness. *Vision Research*, *45*(8), 961–967. https://doi.org/10.1016/j.visres.2004.09.044

Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, *48*(2), 229–240. https://doi.org/10.1111/j.1469-8986.2010.01061.x

Morlet, D., Bouchet, P., & Fischer, C. (2000). Mismatch negativity and N100 monitoring: potential clinical value and methodological advances. *Audiology & Neuro-Otology*, *5*(3–4), 198–206. https://doi.org/13881

Morris, J. S., DeGelder, B., Weiskrantz, L., & Dolan, R. J. (2001). Differential extrageniculostriate and amygdala responses to presentation of emotional faces in a cortically blind field. *Brain*, *124*(6), 1241–1252.

Morris, J. S., Ohman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(4), 1680–1685. https://doi.org/10.1073/pnas.96.4.1680

Morris, J. S., Ohrnan, A., & Dolan, R. J. (1998). Conscious and unconscious emotional learning in the human amygdala. *Nature*, *393*(6684), 467–470. https://doi.org/10.1038/30976

Mumford, D. (1992). On the computational architecture of the neocortex. *Biological Cybernetics*, *66*(3), 241–251. https://doi.org/10.1007/BF00198477

M.W. Spratling. (2008). Predictive coding as a model of biased competition in visual attention. *Vision Research*, *48*(12), 1391–1408. https://doi.org/10.1016/j.visres.2008.03.009

Näätänen, R. (1992). *Attention and brain function*. Psychology Press.

Näätänen, R. (1995). The mismatch negativity: a powerful tool for cognitive neuroscience. *Ear and Hearing*, *16*(1), 6–18.

Näätänen, R., & Escera, C. (2000). Mismatch negativity: clinical and other applications. *Audiology & Neuro-Otology*, *5*(3–4), 105–110. https://doi.org/13874

Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*(4), 313–329. https://doi.org/10.1016/0001-6918(78)90006-9

Näätänen, R., Jacobsen, T., & Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): A review of the evidence. *Psychophysiology*, *42*(1), 25–32. https://doi.org/10.1111/j.1469-8986.2005.00256.x

Näätänen, R., & Michie, P. T. (1979). Early selective-attention effects on the evoked potential: A critical review and reinterpretation. *Biological Psychology*, *8*(2), 81–136. https://doi.org/10.1016/0301-0511(79)90053-X

Näätänen, R., Paavilainen, P., & Reinikainen, K. (1989). Do event-related potentials to infrequent decrements in duration of auditory stimuli demonstrate a memory trace in man? *Neuroscience Letters*, *107*(1–3), 347–352. https://doi.org/10.1016/0304-3940(89)90844-6

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Näätänen, R., Paavilainen, P., Titinen, H., Jiang, D., & Alho, K. (1993). Attention and mismatch negativity. *Psychophysiology*, *30*(5), 436–450. https://doi.org/10.1111/j.1469-8986.1993.tb02067.x

Näätänen, R., Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch negativity (MMN): towards the optimal paradigm. *Clinical Neurophysiology*, *115*(1), 140–144. https://doi.org/10.1016/j.clinph.2003.04.001

Naccache, L., & Dehaene, S. (2001). The Priming Method: Imaging Unconscious Repetition Priming Reveals an Abstract Representation of Number in the Parietal Lobes. *Cerebral Cortex*, *11*(10), 966–974. https://doi.org/10.1093/cercor/11.10.966

N'Diaye, K., Ragot, R., Garnero, L., & Pouthas, V. (2004). What is common to brain activity evoked by the perception of visual and auditory filled durations? A study with MEG and EEG co-recordings. *Cognitive Brain Research*, *21*(2), 250–268. https://doi.org/10.1016/j.cogbrainres.2004.04.006

Neisser, U. (1967). Cognitive psychology. Appleton-Century-Crofts.[aAC] Nelson, K.(2003) Self and social functions: Individual autobiographical memory and collective narrative. *Memory*, *11*(2), 12536.

Nemeth, D., Janacsek, K., Csifcsak, G., Szvoboda, G., Howard Jr, J. H., & Howard, D. V. (2011). Interference between sentence processing and probabilistic implicit sequence learning. *PLoS One*, *6*(3), e17577.

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, *15*(1), 1–25. https://doi.org/10.1002/hbm.1058

Nissen, M. J. (1977). Stimulus intensity and information processing. *Perception & Psychophysics*, *22*(4), 338–352.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32. https://doi.org/10.1016/0010-0285(87)90002-8

NITTONO, H. (2005). Missing-stimulus potentials associated with a disruption of human-computer interaction. *Psychologia*, *48*(2), 93–101.

Novak, G., Ritter, W., & Vaughan, H. G. (1992). Mismatch Detection and the Latency of Temporal Judgments. *Psychophysiology*, *29*(4), 398–411. https://doi.org/10.1111/j.1469-8986.1992.tb01713.x

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*(4), 608–631. https://doi.org/10.1037/0033-295X.101.4.608

Okazaki, S., Kanoh, S., Takaura, K., Tsukada, M., & Oka, K. (2006). Change detection and difference detection of tone duration discrimination. *Neuroreport*, *17*(4), 395–399.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2010). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, e156869. https://doi.org/10.1155/2011/156869

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. MIT press.

Palmer, T. D., & Ramsey, A. K. (2012a). The function of consciousness in multisensory integration. *Cognition*, *125*(3), 353–364. https://doi.org/10.1016/j.cognition.2012.08.003

Palmer, T. D., & Ramsey, A. K. (2012b). The function of consciousness in multisensory integration. *Cognition*, *125*(3), 353–364. https://doi.org/10.1016/j.cognition.2012.08.003

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and SLA. *Implicit and Explicit Learning of Languages*, *393*, 419.

Peigneux, P., Laureys, S., Fuchs, S., Destrebecqz, A., Collette, F., Delbeuck, X., … Degueldre, C. (2003). Learned material content and acquisition level modulate cerebral reactivation during posttraining rapid-eye-movements sleep. *Neuroimage*, *20*(1), 125–134.

Peigneux, P., Maquet, P., Meulemans, T., Destrebecqz, A., Laureys, S., Degueldre, C., … Franck, G. (2000). Striatum forever, despite sequence learning variability: a random effect analysis of PET data. *Human Brain Mapping*, *10*(4), 179–194.

Penney, T. B., Gibbon, J., & Meck, W. H. (2000). Differential effects of auditory and visual signals on clock speed and temporal memory. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(6), 1770.

Pinto, Y., Gaal, S. van, Lange, F. P. de, Lamme, V. A. F., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *Journal of Vision*, *15*(8), 13–13. https://doi.org/10.1167/15.8.13

Pizlo, Z. (2001). Perception viewed as an inverse problem. *Vision Research*, *41*(24), 3145–3161. https://doi.org/10.1016/S0042-6989(01)00173-0

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Rammsayer, T., & Ulrich, R. (2005). No evidence for qualitative differences in the processing of short and long temporal intervals. *Acta Psychologica*, *120*(2), 141–171. https://doi.org/10.1016/j.actpsy.2005.03.005

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855–863. https://doi.org/10.1016/S0022-5371(67)80149-X

Regan, D. (1989). Human brain electrophysiology: evoked potentials and evoked magnetic fields in science and medicine.

Reithler, J., van Mier, H. I., & Goebel, R. (2010). Continuous motor sequence learning: Cortical efficiency gains accompanied by striatal functional reorganization. *NeuroImage*, *52*(1), 263–276. https://doi.org/10.1016/j.neuroimage.2010.03.073

Rock, I. (1983). The logic of perception.

Roeber, U., Widmann, A., & Schröger, E. (2003). Auditory distraction by duration and location deviants: a behavioral and event-related potential study. *Cognitive Brain Research*, *17*(2), 347–357. https://doi.org/10.1016/S0926-6410(03)00136-8

Roger, C., Hasbroucq, T., Rabat, A., Vidal, F., & Burle, B. (2009). Neurophysics of temporal discrimination in the rat: A mismatch negativity study. *Psychophysiology*, *46*(5), 1028–1032. https://doi.org/10.1111/j.1469-8986.2009.00840.x

Romano Bergstrom, J. C., Howard, J. H., & Howard, D. V. (2012). Enhanced implicit sequence learning in college-age video game players and musicians. *Applied Cognitive Psychology*, *26*(1), 91–96.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Sams, M., Paavilainen, P., Alho, K., & Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *62*(6), 437–448. https://doi.org/10.1016/0168-5597(85)90054-1

Schröger, E., & Wolff, C. (1998). Attentional orienting and reorienting is indicated by human event-related brain potentials. *Neuroreport*, *9*(15), 3355–3358.

Sedley, W., Gander, P. E., Kumar, S., Kovach, C. K., Oya, H., Kawasaki, H., … Griffiths, T. D. (2016). Neural signatures of perceptual inference. *ELife*, *5*, e11476. https://doi.org/10.7554/eLife.11476

Seitz, A. R., & Dinse, H. R. (2007). A common framework for perceptual learning. *Current Opinion in Neurobiology*, *17*(2), 148–153. https://doi.org/10.1016/j.conb.2007.02.004

Seriès, P., & Seitz, A. R. (2013). Learning what to expect (in visual perception). *Frontiers in Human Neuroscience*, *7*, 668. https://doi.org/10.3389/fnhum.2013.00668

Seth, A. K. (2014). A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive Neuroscience*, 1–22. https://doi.org/10.1080/17588928.2013.877880

Seth, A. K., Suzuki, K., & Critchley, H. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Consciousness Research*, *2*, 395. https://doi.org/10.3389/fpsyg.2011.00395

Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends in Cognitive Sciences*, *14*(9), 425–432. https://doi.org/10.1016/j.tics.2010.07.001

Sherman, M. T., Seth, A. K., Barrett, A. B., & Kanai, R. (2015). Prior expectations facilitate metacognition for perceptual decision. *Consciousness and Cognition*, *35*, 53–65. https://doi.org/10.1016/j.concog.2015.04.015

Shi, Z., Church, R. M., & Meck, W. H. (2013). Bayesian optimization of time perception. *Trends in Cognitive Sciences*, *17*(11), 556–564. https://doi.org/10.1016/j.tics.2013.09.009

Shih, L. Y. L., Kuo, W.-J., Yeh, T.-C., Tzeng, O. J. L., & Hsieh, J.-C. (2009). Common neural mechanisms for explicit timing in the sub-second range: *NeuroReport*, *20*(10), 897–901. https://doi.org/10.1097/WNR.0b013e3283270b6e

Simson, R., Vaughan Jr., H. G., & Walter, R. (1976). The scalp topography of potentials associated with missing visual or auditory stimuli. *Electroencephalography and Clinical Neurophysiology*, *40*(1), 33–42. https://doi.org/10.1016/0013-4694(76)90177-2

Smith, A. M. (1996). Ptolemy's Theory of Visual Perception: An English Translation of the" Optics" with Introduction and Commentary. *Transactions of the American Philosophical Society*, *86*(2), iii-300.

Song, S., Jr, J. H. H., & Howard, D. V. (2008). Perceptual sequence learning in a serial reaction time task. *Experimental Brain Research*, *189*(2), 145–158. https://doi.org/10.1007/s00221-008-1411-z

Spratling, M. W. (2012). Predictive coding as a model of the V1 saliency map hypothesis. *Neural Networks*, *26*(0), 7–28. https://doi.org/10.1016/j.neunet.2011.10.002

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*(4), 387–401. https://doi.org/10.1016/0013-4694(75)90263-1

Steele, C. J., & Penhune, V. B. (2010). Specific increases within global decreases: a functional magnetic resonance imaging investigation of five days of motor sequence learning. *The Journal of Neuroscience*, *30*(24), 8332–8341.

Stefanics, G., Kremláček, J., & Czigler, I. (2014). Visual mismatch negativity: a predictive coding view. *Frontiers in Human Neuroscience*, *8*, 666. https://doi.org/10.3389/fnhum.2014.00666

Stickgold, R., James, L., & Hobson, J. A. (2000). Visual discrimination learning requires sleep after training. *Nature Neuroscience*, *3*(12), 1237–1238. https://doi.org/10.1038/81756

Stickgold, R., Scott, L., Rittenhouse, C., & Hobson, J. A. (1999). Sleep-Induced Changes in Associative Memory. *Journal of Cognitive Neuroscience*, *11*(2), 182–193. https://doi.org/10.1162/089892999563319

Stickgold, R., & Walker, M. P. (2007). Sleep-dependent memory consolidation and reconsolidation. *Sleep Medicine*, *8*(4), 331–343. https://doi.org/10.1016/j.sleep.2007.03.011

Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, *9*(4), 578–585. https://doi.org/10.1038/nn1669

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409. https://doi.org/10.1016/j.tics.2009.06.003

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), 1004–1006. https://doi.org/10.1038/nn.2163

Sussman, E. (2007). A New View on the MMN and Attention Debate. *Journal of Psychophysiology*, *21*(3), 164–175. https://doi.org/10.1027/0269-8803.21.34.164

Takahashi, K., & Watanabe, K. (2012). Short-term memory for event duration: Modality specificity and goal dependency. *Attention, Perception, & Psychophysics*, *74*(8), 1623–1631. https://doi.org/10.3758/s13414-012-0347-3

Takegata, R., Tervaniemi, M., Alku, P., Ylinen, S., & Näätänen, R. (2008). Parameter-specific modulation of the mismatch negativity to duration decrement and increment: Evidence for asymmetric processes. *Clinical Neurophysiology*, *119*(7), 1515–1523. https://doi.org/10.1016/j.clinph.2008.03.025

Team, R. C. (2014). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013*. ISBN 3-900051-07-0.

Teki, S., Grube, M., & Griffiths, T. D. (2012). A unified model of time perception accounts for duration-based and beat-based timing mechanisms. *Frontiers in Integrative Neuroscience*, *5*, 90. https://doi.org/10.3389/fnint.2011.00090

Teki, S., Grube, M., Kumar, S., & Griffiths, T. D. (2011). Distinct Neural Substrates of Duration-Based and Beat-Based Auditory Timing. *The Journal of Neuroscience*, *31*(10), 3805–3812. https://doi.org/10.1523/JNEUROSCI.5561-10.2011

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to Grow a Mind: Statistics, Structure, and Abstraction. *Science*, *331*(6022), 1279–1285. https://doi.org/10.1126/science.1192788

Tervo, D. G. R., Tenenbaum, J. B., & Gershman, S. J. (2016). Toward the neural implementation of structure learning. *Current Opinion in Neurobiology*, *37*, 99–105. https://doi.org/10.1016/j.conb.2016.01.014

Tiitinen, H., May, P., Reinikainen, K., & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature*, *372*(6501), 90–92. https://doi.org/10.1038/372090a0

Töllner, T., Zehetleitner, M., Gramann, K., & Müller, H. J. (2011). Stimulus Saliency Modulates Pre-Attentive Processing Speed in Human Visual Cortex. *PLOS ONE*, *6*(1), e16276. https://doi.org/10.1371/journal.pone.0016276

Tse, C.-Y., & Penney, T. B. (2006). Preattentive timing of empty intervals is from marker offset to onset. *Psychophysiology*, *43*(2), 172–179. https://doi.org/10.1111/j.1469-8986.2006.389.x

van Beers, R. J., Sittig, A. C., & van Der Gon, J. J. D. (1999). Integration of proprioceptive and visual position-information: An experimentally supported model. *Journal of Neurophysiology*, *81*(3), 1355–1364.

van Turennout, M., Ellmore, T., & Martin, A. (2000). Long-lasting cortical plasticity in the object naming system. *Nature Neuroscience*, *3*(12), 1329–1334. https://doi.org/10.1038/81873

van Wassenhove, V., Buonomano, D. V., Shimojo, S., & Shams, L. (2008). Distortions of Subjective Time Perception Within and Across Senses. *PLoS ONE*, *3*(1), e1437. https://doi.org/10.1371/journal.pone.0001437

VanRullen, R., & Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences*, *7*(5), 207–213. https://doi.org/10.1016/S1364-6613(03)00095-0

VanRullen, R., & Macdonald, J. S. P. (2012). Perceptual Echoes at 10 Hz in the Human Brain. *Current Biology*, *22*(11), 995–999. https://doi.org/10.1016/j.cub.2012.03.050

Vapnik, V. (2013). *The Nature of Statistical Learning Theory*. Springer Science & Business Media.

Vogels, R. (2010). Mechanisms of Visual Perceptual Learning in Macaque Visual Cortex. *Topics in Cognitive Science*, *2*(2), 239–250. https://doi.org/10.1111/j.1756-8765.2009.01051.x

von Kriegstein, K., & Giraud, A.-L. (2006). Implicit Multisensory Associations Influence Voice Recognition. *PLoS Biol*, *4*(10), e326. https://doi.org/10.1371/journal.pbio.0040326

Vuust, P., Brattico, E., Glerean, E., Seppänen, M., Pakarinen, S., Tervaniemi, M., & Näätänen, R. (2011). New fast mismatch negativity paradigm for determining the neural prerequisites for musical ability. *Cortex*, *47*(9), 1091–1098. https://doi.org/10.1016/j.cortex.2011.04.026

Vuust, P., Ostergaard, L., Pallesen, K. J., Bailey, C., & Roepstorff, A. (2009). Predictive coding of music – Brain responses to rhythmic incongruity. *Cortex*, *45*(1), 80–92. https://doi.org/10.1016/j.cortex.2008.05.014

Wacongne, C., Changeux, J.-P., & Dehaene, S. (2012). A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. *Journal of Neuroscience*, *32*(11), 3665–3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(51), 20754–9. https://doi.org/10.1073/pnas.1117807108

Walker, M. P., & Stickgold, R. (2006). Sleep, Memory, and Plasticity. *Annual Review of Psychology*, *57*(1), 139–166. https://doi.org/10.1146/annurev.psych.56.091103.070307

Wassenhove, V. van. (2009). Minding time in an amodal representational space. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1525), 1815–1830. https://doi.org/10.1098/rstb.2009.0023

Wearden, J. H., Todd, N. P. M., & Jones, L. A. (2006). When do auditory/visual differences in duration judgements occur? *The Quarterly Journal of Experimental Psychology*, *59*(10), 1709–1724. https://doi.org/10.1080/17470210500314729

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*(6), 598–604. https://doi.org/10.1038/nn0602-858

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee Michael, B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience*, *18*(1), 411–418.

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, *8*(2), 227–233. https://doi.org/10.1016/S0959-4388(98)80144-X

Wijnen, V. J. M., van Boxtel, G. J. M., Eilander, H. J., & de Gelder, B. (2007). Mismatch negativity predicts recovery from the vegetative state. *Clinical Neurophysiology*, *118*(3), 597–605. https://doi.org/10.1016/j.clinph.2006.11.020

Williams, D., Phillips, G., & Sekuler, R. (1986). Hysteresis in the perception of motion direction as evidence for neural cooperativity. *Nature*, *324*(6094), 253–255. https://doi.org/10.1038/324253a0

Williams, M. A., Morris, A. P., McGlone, F., Abbott, D. F., & Mattingley, J. B. (2004). Amygdala Responses to Fearful and Happy Facial Expressions under Conditions of Binocular Suppression. *Journal of Neuroscience*, *24*(12), 2898–2904. https://doi.org/10.1523/JNEUROSCI.4977-03.2004

Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1047.

Winkler, I., & Czigler, I. (2012). Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations. *International Journal of Psychophysiology*, *83*(2), 132–143. https://doi.org/10.1016/j.ijpsycho.2011.10.001

Winkler, I., Haufe, S., & Tangermann, M. (2011). Automatic Classification of Artifactual ICA-Components for Artifact Removal in EEG Signals. *Behavioral and Brain Functions*, *7*, 30. https://doi.org/10.1186/1744-9081-7-30

Wu, C.-T., Busch, N. a, Fabre-Thorpe, M., & VanRullen, R. (2009). The temporal interplay between conscious and unconscious perceptual streams. *Current Biology : CB*, *19*(23), 2003–7. https://doi.org/10.1016/j.cub.2009.10.017

Wuerger, S., Shapley, R., & Rubin, N. (1996). On the visually perceived direction of motion by Hans Wallach: 60 years later. *Perception-London*, *25*(11), 1317–1368.

Xu, S., Jiang, W., Poo, M., & Dan, Y. (2012). Activity recall in a visual cortical ensemble. *Nature Neuroscience*, *15*(3), 449–455. https://doi.org/10.1038/nn.3036

Yang, T., & Maunsell, J. H. R. (2004). The Effect of Perceptual Learning on Neuronal Responses in Monkey Visual Area V4. *The Journal of Neuroscience*, *24*(7), 1617–1626. https://doi.org/10.1523/JNEUROSCI.4442-03.2004

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The Neural Basis of Error Detection: Conflict Monitoring and the Error-Related Negativity. *Psychological Review*, *111*(4), 931–959. https://doi.org/10.1037/0033-295X.111.4.931

Yotsumoto, Y., Sasaki, Y., Chan, P., Vasios, C. E., Bonmassar, G., Ito, N., … Watanabe, T. (2009). Location-Specific Cortical Activation Changes during Sleep after Training for Perceptual Learning. *Current Biology*, *19*(15), 1278–1282. https://doi.org/10.1016/j.cub.2009.06.011

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, *10*(7), 301–308. https://doi.org/10.1016/j.tics.2006.05.002

Zhou, W., Jiang, Y., He, S., & Chen, D. (2010). Olfaction Modulates Visual Perception in Binocular Rivalry. *Current Biology*, *20*(15), 1356–1358. https://doi.org/10.1016/j.cub.2010.05.059